

# ニューラルネットワークによる薬物の構造活性相関に関する研究

著者	青山 智夫
学位授与機関	Tohoku University
学位授与番号	321
URL	<a href="http://hdl.handle.net/10097/46020">http://hdl.handle.net/10097/46020</a>

ニューラルネットワークによる薬物の

構造活性相関に関する研究

青山智夫

# ニューラルネットワークによる薬物の構造活性相関に関する研究

## 目 次

緒 言	3
第1章 ニューラルネットワークの概要	
1.1 ニューロコンピュータの歴史	5
1.2 形式ニューロンの動作とニューラルネットワークの方式	7
第2章 パーセプトロン型ニューラルネットワークの理論	
2.1 用語の定義及び基本的動作	12
2.2 バックプロパゲーション法	17
2.3 パーセプトロン型ネットワークの動作と従来方法との関係	21
2.4 パーセプトロン型ネットワークの層数	23
第3章 パーセプトロン型ネットワークの分類及び fitting 特性	
3.1 分類	25
3.2 Fitting	39
3.3 ニューラルネットワーク法の問題点	43
3.4 leave-one-out法	49
第4章 ニューラルネットワーク法の活性相関研究への応用	
4.1 等級活性分類への応用	51
4.2 定量的構造活性相関研究への応用	60

第5章 ニューラルネットワーク法の活性相関研究の展開	
5.1 相関指数	75
5.2 再構築学習法	86
結 語	99
文 献	103
謝 辞	107
付 録	108



## 緒 言

望ましい医薬品を効率良く開発するため、また薬物の生理活性の本質を理解するため、薬物の生理活性の強度と化学構造との関係を調べる研究が行われて来た。この問題に関しては従来から統計学および多変量解析が用いられていたが、最近新しい考え方、例えば学習機械法、シンプレックス法、最近隣法、Simple Classification Algorithm法、Adaptive Least-Squares (適応最小2乗) 法などが提唱されている。これらの方法は情報科学の観点からはパターン認識技術の一つである。即ち、化合物のスペクトルデータ、化学構造データを一種のパターンとみなし、それらと生物活性強度との相関関係を解析する。

脳における神経細胞の情報処理についての研究が進展し、それに基づく学習機能、分類機能、及び予測機能を有するいくつかの神経回路網モデルが提出された。これらモデルをニューラルネットワークと呼ぶようになった。なお、ニューロコンピュータはニューラルネットワークを情報処理の基本素子として具備する計算装置である。

当初ニューラルネットワークは脳のパターン認識機構との関連についての研究であつた。それに平行し、ニューラルネットワークの情報処理法の数学的な性質が研究された結果、非線形的な統計処理動作が注目されるようになった。今日統計学は因果関係を手続き的な関係で示すことが困難である事象に適用されている。特に重回帰分析法は有効である。しかし線形の因果関係から外れるにしたがい、また因果関係に内在する非線形関係が不明の場合、重回帰法の適用は困難である。

ニューラルネットワークはその構成素子であるニューロンの固有動作に基づいて動作する。その固有動作はネットワークの中で、外部より提示された事例に基づいて積算され、総じて原因と結果の非線形的な因果関係をネットワーク内に形成する。そこでは手続き的な処理を記述する必要がないため、因果関係が従来の

統計的手段で解析困難な場合に、ニューラルネットワークによる解析の効果が期待できる。このようなニューラルネットワークの特徴は生物が関係するデータの処理、未知量の予測に適すると考えられる。

ニューラルネットワークには種々の結合方式が可能である。これらのうち、最も単純でかつその数学的な性質が明らかになっているものは階層型（パーセプトロン型）のニューラルネットワークである。パーセプトロン型ニューラルネットワークは複数のニューロン層から構成され、入力された情報はこれらの層を順次伝播するうちに変形され、情報を構成する特徴が抽出される。この型のネットワークの特徴は他の型に比べ効率的な学習方法が確立されていることである。すなわち、入力した情報を目的とする表現形に変形する能力を高速学習によって獲得することができる。

本論文はパーセプトロン型ニューラルネットワークのパターン認識動作を解析し、薬物の構造活性相関研究へ適用し、さらに目的にそった展開を試みたものである。

## 第1章 ニューラルネットワークの概要

本章ではニューロコンピュータの歴史およびニューラルネットワークの動作と方式を概説する。

### 1.1 ニューロコンピュータの歴史

ニューラルネットワークはニューロコンピュータを構成する演算素子であり、神経細胞（ニューロン）の情報伝達機構を模した情報処理方式である。ここにその発展史の概略を述べる。

1943年 McCulloch と Pitts はニューロンの動作に関して次のようなモデルを提案した。ニューロンは興奮すると出力側の軸索にパルスを出し、興奮していないときはほとんど出さない。一つのニューロンは他の多くのニューロンから受け取った刺激の総和がその細胞ごとに決められた閾値を越えると興奮し、別のニューロンにパルスを送る。<sup>1)</sup> こうして記述されるモデルは形式ニューロンと呼ばれている。

1949年 Hebb はニューロン間の結合部（シナプス）は、刺激が伝達される程、その結合強度が増加し、さらに刺激が伝えやすくなり、これが神経回路に可塑性をもたらして、認識や記憶を形成していると仮定した。<sup>2)</sup> この仮説はHebbのシナプス強化則（Hebbian rule）と呼ばれている。Hebbian ruleを形式ニューロンモデルに適用し、ニューロン間の結合方式を吟味したものが今日のニューラルネットワークの基本原理である。

1962年 Rosenblatt はこれらの原理に基づいてつぎのようなパターンを識別する学習機械パーセプトロンをつくった。<sup>3)</sup> パーセプトロンは非常に多くの感受素子からなる受容層（sensory unit, S層）、複数のモデルニューロンからなる連合層（association unit, A層）および1個のモデルニューロンからなる出力

層または反応層 (response unit, R層) で構成されている。各層の細胞は0または1の値をとるものとする。あるカテゴリーに属する図形を受容層に与えて出力細胞を興奮させ (値1を取らせる), そのカテゴリーに属さない図形を与えたとき出力細胞の興奮を抑える (値0を取らせる) ようなデータを与えていくと, 出力細胞はそのカテゴリーに属したものに対してのみ興奮するようになる。すなわち, この機械は図形を識別するようになる。Hebbian rule はシナプス結合が強化されることのみを主張するが, 上記のパーセプトロンでは, 出力層の細胞が興奮すべきでないときに興奮したときはシナプス間の結合を弱める働きも組み込まれている。

この機械の提案に対しては大きな反響があり, その後多くの研究者らにより研究が展開された。1962年 Blockは, Hebbian rule にシナプス結合を弱めるという逆の働きを加えた場合, 学習が正しい分類に必ず収束するという, いわゆるパーセプトロンの収束定理を得た。<sup>4)</sup> この定理はパーセプトロン研究に拍車をかけることになった。

1969年 Minsky と Papert は特定の構造のパーセプトロンを用い, 学習収束定理, 収束速度, およびシナプス結合の強度の変動等について検討し, 実用上の限界を示した。<sup>5)</sup> 人工知能の分野での彼らの影響力は強く, この発表以降パーセプトロン研究は急速に下火になった。しかし, 少数の研究者によって地道な研究は続けられ, ニューラルネットワークに関する基礎的な知識が確実に蓄積されてきた。

1969年中野は連想記憶の機能をもつニューラルネットワークであるアソシアトロンを提案し,<sup>6)</sup> 同様なネットワークは上坂と尾関 (1972年),<sup>7)</sup> Anderson (1972年),<sup>8)</sup> や Kohonen (1972年) からも提案された。<sup>9)</sup> また, 文字に対して自己組織化機能を持ち, それによって教師なし学習を行う福島のコグニトロン (1975年) も発表された。<sup>10)</sup> ニューラルネットワークの性質の理論的研究では,

甘利,<sup>11)</sup> Malsburg の研究がある。<sup>12)</sup>

1980年代になって、米国を中心にニューラルネットワーク研究は新たな展開を見せた。Rumelhart, Hinton, Sejnowski らはPDP (Parallel Distributed Processing) と称するグループをつくり、ネットワーク構造を有する並列分散情報処理装置の研究を始めた。その研究の目指すものをニューロコンピュータと名付けた。この間2つの有名なネットワーク構造が提案されている。一つはホップフィールドモデル (1982年),<sup>13)</sup> 他はボルツマンマシンモデル (1985年) である。<sup>14)</sup>

1986年 Rumelhart ら、PDPグループはバックプロパゲーション (back-propagation, 逆伝播) 法と呼ばれるパーセプトロン型ニューラルネットワークに対する画期的な学習方法を提案した。<sup>15)</sup> それは出力層のニューロンに教師信号を提示し正解と食い違った場合、最急降下法という方法に基づきニューロン間の結合の強度を修正し、最終的に入出力関係を満たすようなネットワークとする方法である。この方法により、Minsky と Papert によって指摘されたパーセプトロンの困難の一部は解決された。Rosenberg と Sejnowski は、英文を読むことをこの方法で学習させ NETtalk と名付けられたシステムをつくった。<sup>16)</sup> この成果により、バックプロパゲーション法の応用研究が開始された。

## 1. 2 形式ニューロンの動作とニューラルネットワークの方式

この節は形式ニューロンの動作、ニューロンの結合方式について解説する。

### 1. 2. 1 形式ニューロンとそのシミュレーション

ニューロンの構造及び動作は単純ではないが、その機能をごく単純化すると、  
(1) シナプス前細胞の出力信号がシナプス結合をとおしてシナプス後細胞の膜電位に変化を与え、(2) 多くのシナプスの出力信号の総和によって膜電位が決

まり、そして(3)膜電位がある閾値を越えると、シナプス後細胞は興奮し軸策に出力信号を与えるという機能に集約できる。このような機能のみを有するニューロンを形式ニューロンと呼ぶ。

形式ニューロンが興奮し出力軸策に出力信号を与える状態を1, そうでない状態を0とすると, 形式ニューロンの動作は次式で示すことができる。

$$y_j = \sum_i w_{ij} x_i - \theta \quad 1-1$$

$$z = f(y_j) \quad 1-2$$

$$f(y_j) = \{1 + \text{sgn}(y_j)\} / 2 \quad 1-3$$

ここで,  $\theta$  は閾値,  $x_i$  はシナプス前細胞の出力値(0または1)である。 $w_{ij}$  はシナプス荷重あるいは結合荷重(行列)とよばれ, シナプスの結合の強度をあらわす。 $f$  はニューロンの動作を表す関数である。以下動作関数と言う。

ニューロンの動作を数式処理するには, 微分出来る方が便利であり, かつ生体で実現可能な関数として, シグモイド関数が階段関数の代わりに使われている。本論文でもシグモイド関数を使用する。

$$f(y) = 1 / \{1 + \exp(-\alpha y)\} \quad 1-4$$

## 1. 2. 2 形式ニューロンの動作

いま式1-1 においてそのニューロン固有の  $w = \{w_1, w_2, \dots\}$  が決定されているものとする。式1-1 の左辺を0とおくと, ベクトル  $x = \{x_1, x_2, \dots\}$  の張る空間で,  $w$  と閾値  $\theta$  によって勾配と位置が定まる超平面となる。この空間内ではベクトル  $x$  は点で表され, 超平面を境にして式1-2 の値は0/1を与える。従って, 形式ニューロンは結合しているニューロンの出力をそのニューロン独自の判断基準( $w$ と $\theta$ )を基に認識・分類する作用を有する。

### 1. 2. 3 ニューロンの結合方式

形式ニューロンの結合によってニューラルネットワークが形成される。ネットワークは大別して階層型と相互結合型とがある。

図1-1a に示すように、階層型ネットワークは入力層、単数または複数の中間層（かくれ層とも呼ばれる）、および出力層から成り、情報は入力層から出力層に向かって伝播する。1962年 Rosenblatt がはじめて提案した形であるので、本論文ではパーセプトロン型（ニューラル）ネットワークと呼ぶ。パーセプトロン型ネットワークはバックプロパゲーション法という学習方式が確立されているため、<sup>15)</sup> もっとも多く利用されている構造である。著者の研究もこの方式を基礎とした。

相互結合型のネットワークは階層型のネットワークのような特別な構造は持たず、ニューロンは相互に結合され情報も双方向に伝達される。相互結合型のネットワークの代表的な例として、ホップフィールド型ネットワークとボルツマンマシンと呼ばれるネットワークがある。

ホップフィールド型ネットワーク（図1-1b）は 1982年 Hopfield によって提案された、<sup>13)</sup> 物質の電子スピンの相互作用をモデルとしたネットワークである。結合は平面的に隣接するニューロン間を相互結合させた構造を有し、ニューロンの動作は非同期的である。このネットワークの特徴はアナログ電子回路で容易に実現でき、収束が高速であり、また連想記憶装置に応用できる点である。しかしネットワークに多くの結合の停留値（ネットワークエネルギーの極小値という）が存在し、最小値への到達が困難である。

一方、ボルツマンマシン（図1-1c）は、1985年 Hinton らによって提案されたネットワークである。<sup>14)</sup> このネットワークはすべてのニューロンが相互結合した構造を有し外界と直接情報をやりとりしない隠れ層がある。また学習則が定義される。

ホップフィールドネットワークでは、ネットワークの状態はエネルギーを必ず減少させるような方向に変化するが、ボルツマンマシンではある確率でエネルギーを増加するような変化も許す。この拡張によって、ホップフィールドネットワークでは困難であった系の状態を最小値へ収束させることが可能になった。ただし、計算に要する時間は膨大である。著者は、このネットワークを構造活性相関の研究対象とはしなかった。



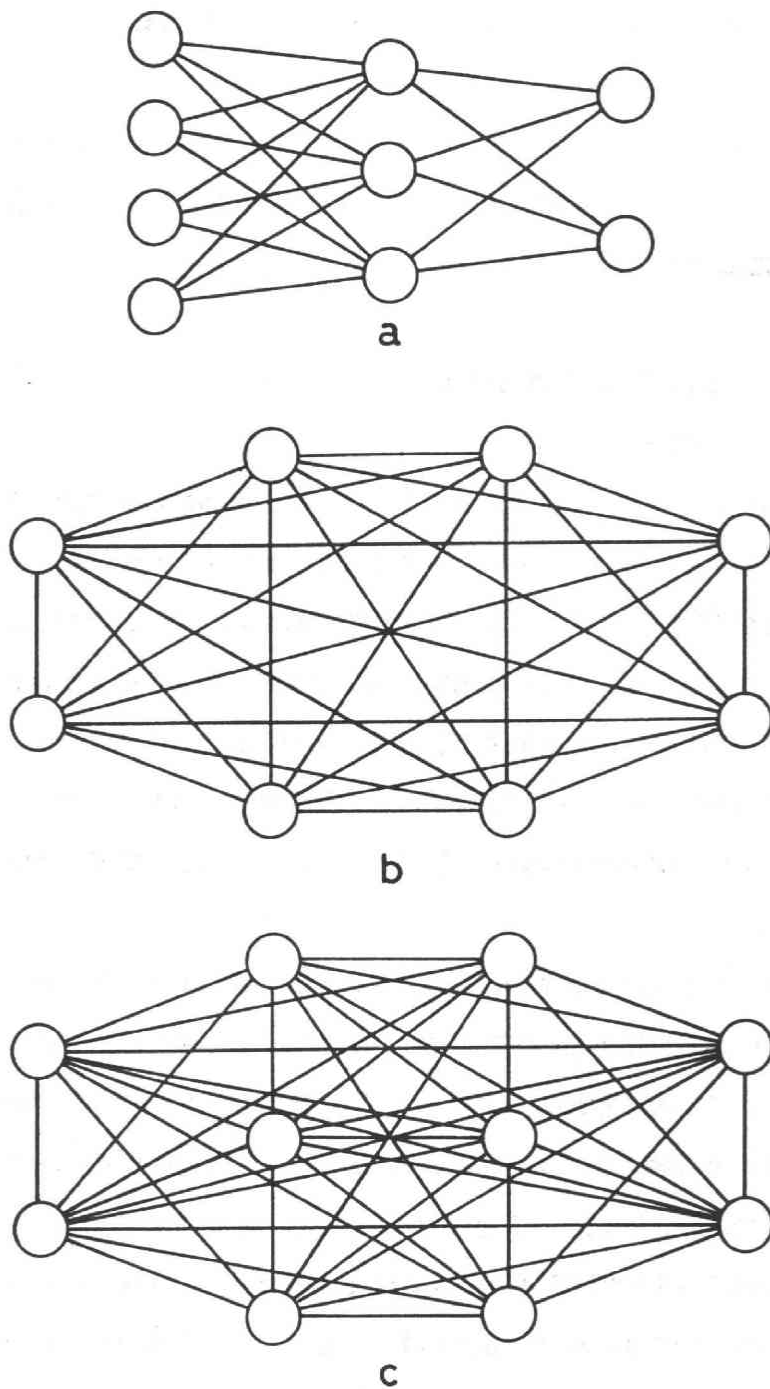


図1- 1.ニューラルネットワークの結合方式

a. パーセプトロン型      b. ホップフィールド型

c. ボルツマンマシン型

## 第2章 パーセプトロン型ニューラルネットワークの理論

本章はパーセプトロン型ニューラルネットワークの基本的動作，及びニューラルネットワークの近似と適応最小2乗（ALS）法及び重回帰分析法との関係について議論する。

### 2.1 用語の定義及び基本的動作

#### 2.1.1 学習

学習とは系が外界からの入力に応じてその系自身の構造を変形（自己組織化）することであると定義できる。生物の神経系においてはニューロン間のシナプス結合の強さを変化させることによって学習が行われるものと理解されている。ニューラルネットワークの理論においては，学習はシナプス間の結合荷重を入力に応じて変化させることを意味する。また，入力と出力のみに注目すれば，学習とは入出力関係を変化させることであり，内部構造までは規定されない。このことは，同じ入出力関係でも環境によってネットワークは異なる構造をとることを意味する。

教師信号の有無により学習は，教師あり学習と教師なし学習とに分類することができる。学習の指針として，ある入力に対しネットワークの望ましい出力が与えられている場合を教師あり学習という。外部から与えられる教師信号を教師データあるいは教師パターンと呼ぶ。教師なし学習は入力信号の統計的性質をネットワークに取り込み，その構造編成に反映させるものである。

薬物の構造活性相関研究は，既知の化合物の構造と生物活性データから未知化合物の生物活性を予測する方法を確立することを主な目的とする。化合物の構造データを入力信号，生物活性データを教師信号とみなせば，教師あり学習を適用することが適切と考えられるため，本研究ではその手法を用いた。

## 2. 1. 2 パターン認識

通常パターンは幾何学的な図形を連想させるが、ここでは拡大された概念を扱う。すなわち、個々のパターンは数値を並べたベクトルで表現されるものと考え、これをパターンベクトルと呼ぶ。

パターンベクトルのうち同じものとすべきパターンの集合をカテゴリーと呼ぶ。 $n$  個の要素を持つパターンベクトル  $p = \{p_1, p_2, \dots, p_n\}$  が与えられたとき、要素を一つ一つの座標に割当てた  $n$  次元の空間（これをパターン空間と呼ぶ）を考える。この空間内では各パターンは点で表現される。同じカテゴリーに属するパターンは相互に近いミンコフスキー距離にある点と考え、図2-1 に示すようにカテゴリーごとに集まって分布すると考える。この時、パターン認識とはパターン空間内の点をカテゴリーごとに分割することである。この分割にはパターン空間内で直線（高次の空間では平面または超平面と呼ぶ）で分離する場合と曲線（曲面または超曲面）で分離する場合とがある。前者を線形分離（分類）また後者を非線形分離（分類）と呼ぶ。

## 2. 1. 3 学習則と収束定理

式1-1 においてベクトル  $x$  をパターンベクトル  $p$  とみなし、1.2.2 節の議論を適用すれば、形式ニューロンは超平面により入力パターンを 2 つのカテゴリーに線形分離する機能を有する。

つぎに、教師信号付きの線形分離可能なパターン分離問題が与えられた場合を考える。カテゴリーに分類する超平面の位置と傾きは、出力が外部からの教師ベクトル  $t$  に適合するように結合荷重  $w$  を調節し、試行錯誤的な反復学習により決定する。この学習則を用いれば、形式ニューロンは有限回の反復学習で正しい認識をすることが証明されている（パーセプトロンの収束定理）。<sup>4)</sup>

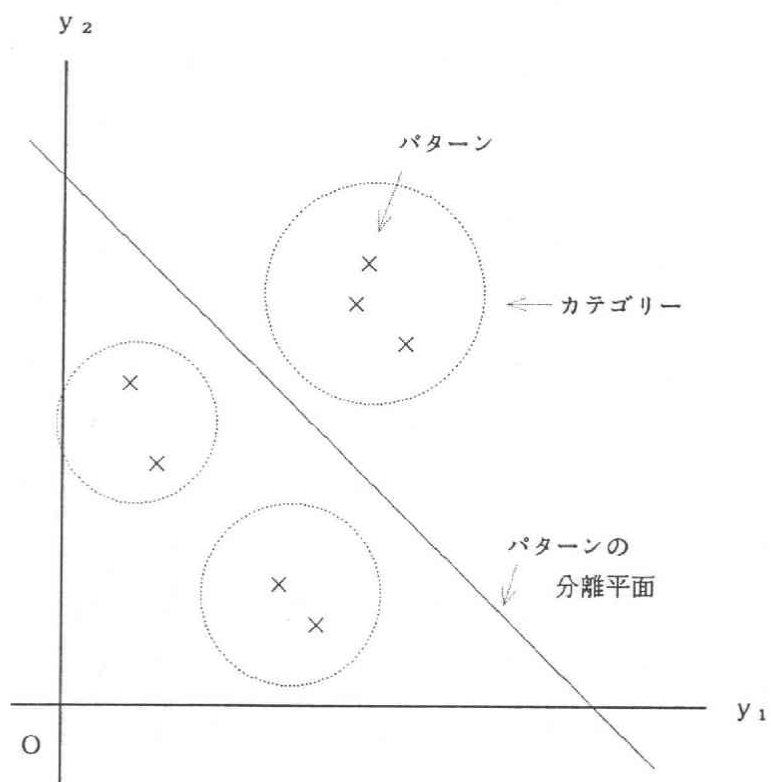


図2-1. パターン空間内のパターンベクトルの分布

#### 2. 1. 4 単純パーセプトロン

線形分離可能な任意のパターン分類問題は有限回の反復学習で達成できるが、現実にはそのまま適用できる問題は少ない。この解決策の一つとして、パターンベクトルを何らかの方法で線形分離可能な形に前処理をする必要がある。これはニューロンの配列を階層構造化することで達成できる。

一例として、図2-2 Rosenblatt によって提出された単純パーセプトロンと呼ばれる3層階層構造を有するネットワークを示す。<sup>3)</sup> A層はS層の各信号を組み合わせ分類し、中間信号を生成してR層へ送り、R層はA層からの信号を組み合わせ非線形分離動作を行う。S層からA層への結合強度がランダムになっており、A層でランダム変換を行う。このランダム変換回路は入力パターンの線形分離性を高める性質がある。分離能力の目安として、k個のn次元パターンベクトルを線形分離可能なように2つのカテゴリーに分ける分け方の個数L(k, n)は次のように与えられることが甘利によって示された。<sup>11)</sup>

$$\lim_{n \rightarrow \infty} \frac{L(k, n)}{2^k} \begin{cases} = 1 & k/n \leq 2 \\ = 0 & k/n > 2 \end{cases} \quad 2-1$$

これはA層の素子数が十分に大きいとき、2n個以下の個数の入力パターンなら線形分離が可能であることを示す。すなわち、単純パーセプトロンの識別限界は2nである。素子数が多ければ非線形分類でも可能であるという単純パーセプトロンの性質は魅力的であるが、無数の素子を必要とする情報処理方式は実用的には意味がない。したがって、限られた数の素子でも有効な学習方式を考える必要がある。

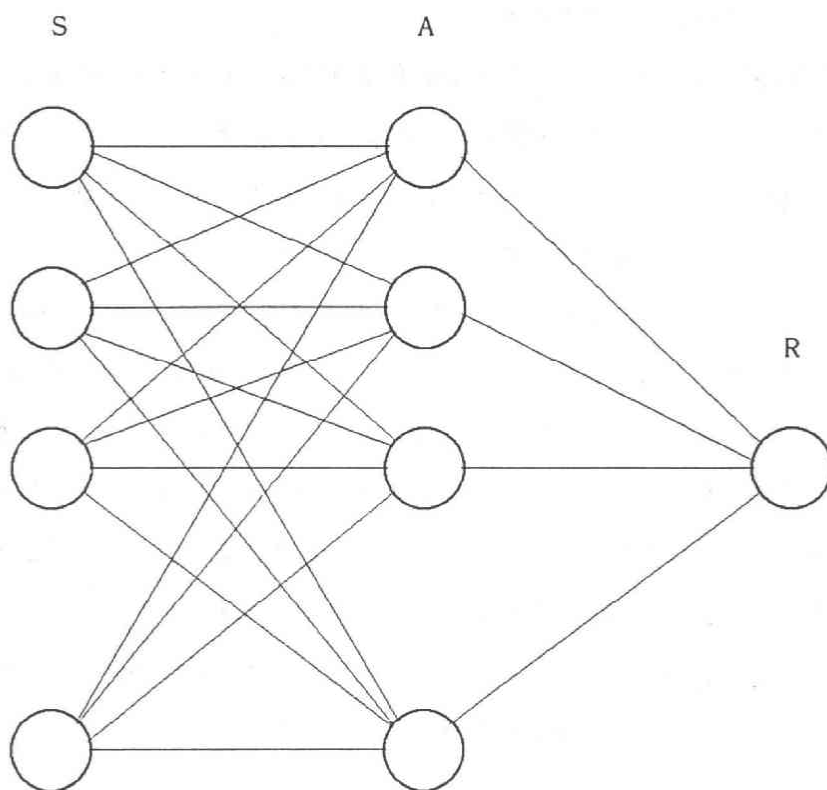


図2-2. 単純パーセプトロン

S : sensory unit

A : association unit

R : unit  
response

## 2. 2 バックプロパゲーション法

単純パーセプトロンのように中間層の結合強度がランダムでは分離のため必要な素子数が多くなる。一方、多層にすることにより、線形分離能力を高めることができるが、処理に多くの時間を要する。そこで、有限の中間層の素子を用い、S, A層間の結合に積極的に学習させ、問題に応じた中間信号を出力させる回路をつくる必要がある。この学習を階層型ネットワークにおいて実現したのが、Rumelhart らによって提出されたバックプロパゲーション法<sup>15)</sup>である。

### 2. 2. 1 デルタルール

階層型ネットワークにおける学習を想定し、形式ニューロン素子の動作を次のように規定する。ある素子  $j$  は前層の素子  $i$  の出力  $x_i$  を入力として受け、結合荷重  $w_{ij}$  を掛けて加えたものの総和を  $y_j$  とする。 $j$  の出力  $O_j$  は  $y_j$  に動作関数  $f$  を施したもので表されるものとし、0~1 間の値をとるものとする。

$$\begin{aligned} O_j &= f(y_j), \\ y_j &= \sum w_{ij} x_i \end{aligned} \quad 2-2$$

ただし、ニューロン  $j$  の閾値は結合荷重に含まれているものとする。

ネットワークがある入力ベクトルに対し出力  $O_j$  を与えたとする。学習を評価するため次のような誤差関数  $E$  を定義する。

$$E = (1/2) \sum (O_j - t_j)^2 \quad 2-3$$

$t_j$  は教師ベクトルの要素である。 $O_j$  は素子間の結合荷重  $w_{ij}$  で決まるため、誤差関数は  $w_{ij}$  の関数となる。 $E$  は  $w_{ij}$  の空間上の超曲面とみなすことができ、この誤差曲面の極小値に達するには、 $w_{ij}$  を  $E$  の  $w_{ij}$  についての偏微分  $DE/Dw_{ij}$  に比例した量 ( $\delta w_{ij}$ ) だけ変化させればよい。

$$\delta w_{ij} = -\epsilon \frac{DE}{Dw_{ij}} \quad 2-4$$

これは誤差曲面上を最も急な傾斜方向に進んで行くことに相当し、この学習を最急降下法という。

式2-3 において、Eを $w_{ij}$ で偏微分する。

$$\frac{DE}{Dw_{ij}} = \sum \frac{DE}{DO_j} \frac{dO_j}{dy_j} \frac{Dy_j}{Dw_{ij}} \quad 2-5$$

ここで、D は偏微分記号を、d は全微分記号を意味する。式2-2 より、

$dO_j/dy_j = f'(y_j)$  ,  $Dy_j/Dw_{ij} = x_i$ であり、式2-3 より、

$DE/DO_j = O_j - t_j$  であるから、式2-4 は、

$$\delta w_{ij} = -\epsilon (O_j - t_j) f'(y_j) x_i \quad 2-6$$

となる。この学習則をデルタルールという。

$f'$  は $f$ がシグモイド関数であるとき、

$$f'(y_j) = f(y_j) \{1 - f(y_j)\} \alpha \quad 2-7$$

である。したがって、このとき $w_{ij}$ の変化量 $\delta w_{ij}$ は、

$$\delta w_{ij} = -\epsilon (O_j - t_j) f(y_j) \{1 - f(y_j)\} \alpha x_i \quad 2-8$$

である。

## 2. 2. 2 バックプロパゲーションアルゴリズム

図2-3 にパーセプトロン型ニューラルネットワークの構造を示す。○印はすべて形式ニューロンで、その動作は式2-2 に示す動作関数に規定されるものとする。入力データはAから入り、Bから出力される。パーセプトロン型ネットワークでは、どの素子も一つ前の層からのみ入力を受け、次の層のみへ出力するものとする。このようなネットワークの $DE/Dy_j$ は出力層より逆向きに計算していくことになる。つまり、出力の誤差を順に前の層へ伝えていくのがバックプロパゲーションである。この考え方にデルタルールを適用し、次のアルゴリズムを得る。



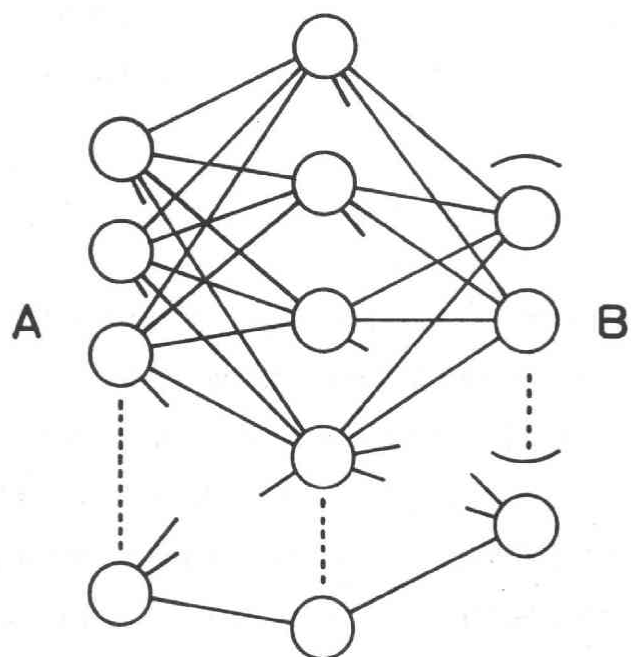


図2-3. パーセプトロン型ニューラルネットワーク

$$\delta w_{ji} = -d_j x_i \varepsilon \quad 2-9$$

$$d_j = (O_j - t_j) f'(y_j) \quad 2-10$$

$$d_j = \{\sum w'_{ji} d'_{ji}\} f'(y_j) \quad 2-11$$

ここで、式2-10 は第2, 3層の修正に、また式2-11 は第1, 2層の修正に用いる。 $w'_{ji}$ および  $d'_{ji}$ は第3層での $w_{ji}$ および $d_j$ である。

ニューロンの動作関数のとり得る値は 0 ~ 1 の範囲であるため、入力データは次式によりスケールする。

$$x_{scaled} = \frac{(q_{max} - q_{min}) x + q_{min} x_{max} - q_{max} x_{min}}{x_{max} - x_{min}} \quad 2-12$$

ここで、 $x$ は入力データ、 $x_{max}$ と $x_{min}$ は入力データの最大値及び最小値、また $q_{max}$ と $q_{min}$ はスケール値の最大および最小値である。

カテゴリーの表現方法はいくつか考えられる。最も単純な一つは、例えば、5段階分類で等級1および5に対しそれぞれ、(0,0,0,0,1) および (1,0,0,0,0)のベクトルで表すことである。この表現の利点は他のクラスの混入の程度がわかる点である。学習は誤差関数E (式2-3) が十分小さくなるまで繰り返し行う。M組の入力と教師ベクトルに対し以上のようなアルゴリズムで学習させると、ニューラルネットワークは入力ベクトルを複数のカテゴリーに分類する能力を持つようになる。スケーリングとカテゴリー表現についての問題点を付録4に示す。

結合荷重行列について次のような特徴がある。同じ入出力関係でも学習時の環境によって、結合荷重行列要素の値は異なる。例えば、3層構造のネットワークにおいて、1, 2層間の荷重が中間層のニューロン $j_0$ に集まるものとする。ニューロン $j_0$ を $j_1$ と $j_2$ に分離して同様の学習を行えば、同じ結果が得られる。そのとき、ニューロンを加える前の荷重 ( $w_{ji}(0)$ ) と加えた後の荷重には、

$$w_{ij}(0) = \lambda_1 w_{ij}(1) + \lambda_2 w_{ij}(2) \quad 2-13$$

の関係がある。荷重値  $w_{ij}(1)$  と  $w_{ij}(2)$  は係数  $\lambda_1$  及び  $\lambda_2$  の値に依存し不定である。

## 2. 3 パーセプトロン型ネットワークの動作と従来方法との関係

薬物の構造活性相関研究では、等級分類を目的とした適応最小 2 乗 (ALS) 法<sup>18)</sup> また定量的構造活性相関には重回帰分析法が用いられる。これらの方法は線形分類及び線形 fitting である。一方、ニューラルネットワークは基本的には非線形動作を行う。本節ではニューラルネットワークの近似と従来の線形的手法との関係を明確にし、それらの方法の特徴を検討する。

### 2. 3. 1 分類動作と適応最小 2 乗 (ALS) 法との関係

ALS 法での識別関数  $L$  は、 $m$  個の記述変数  $X_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$  で表現される。

$$L = XW \quad 2-14$$

識別則は、 $L$  の値が  $a_n < L < a_{n+1}$  のとき、そのグループを  $n$  とする。  $j$  サイクルにおける重み係数  $W(j)$  は、

$$W(j) = (X^T X)^{-1} X^T S(j) \quad 2-15$$

で求める。ここで、 $S$  は補正項  $C$  を用い、

$$\begin{aligned} S(j+1) &= L(j), & (\text{分類が一致したとき}) \\ &= L(j) - C(j) & (\text{分類が一致しないとき}) \end{aligned} \quad 2-16$$

で与えられる。 $W$  の次元が 0 であるので、 $S$  は  $X$  と同じ数学的および物理的性質を有する。ALS 系では、 $X$  と同じ次元を有する量はないので、式 2-15 は適切である。

パーセプトロン型ネットワークにつぎのような構造上および動作上の制限を加

えることにより，ALS法の動作を実現させることができる。

- (1) 2層のネットワークを用いる。
- (2) ニューロンの動作関数として線形関数を用いる。
- (3)  $\theta$  を  $a_n$  に設定する。
- (4)  $w_{ij} = w_{ik}$  ( $j, k$  は第2層の任意のニューロン) とする。
- (5) クラス  $n$  に対し  $n$  個の出力ニューロンを発火させるような教師パターンを与える。

以上の制限により，第2層のニューロン  $i$  への出力値  $z_i$  は，

$$z_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots = w_0 + \sum_j w_{ij} x_{ij} \quad 2-17$$

となる。図2-4 に示すように識別関数  $L_i$  を，

$$L_i = z_i - \theta_i \quad 2-18$$

と定義することにより，ALS法の要件は満たされる。

### 2. 3. 2 MR型ネットワークと重回帰分析法との関係

出力層のニューロンの種類を1種にし，その動作関数を，

$$f(y_i) = y_i \quad 2-19$$

とし，0～1の連続値をとらせることにより，ネットワークは重回帰分析法に似た動作をする。この型のネットワークをMR型（ニューラル）ネットワーク<sup>19)</sup>と呼ぶ。

MR型ネットワークの動作と重回帰分析法との関係を示す。議論を簡単にするため，3層構造のMR型ネットワークについて説明する。ニューロンの閾値  $\theta_i$  は荷重行列の中に含めることができるから，

$$y = w x \quad 2-20$$

と書ける。ここで， $w$  は重み行列， $x$  は入力ベクトルである。すべてのニューロ

ンの動作関数が式2-19（即ち線形関数）によるとすれば，出力ベクトル  $z$  は，  
1，2層間，2，3層間の重み行列を  $w_1, w_2$  として，

$$z = (w_1 w_2) x = w x \quad 2-21$$

と表わすことができる。一方，重回帰分析法は次の一次方程式の最適係数  $a_i$  及び  $b_{ij}$  を求めることである。

$$z_i = a_i + \sum_j b_{ij} x_j = \sum_j b'_{ij} x'_j \quad 2-22$$

ただし，

$$x'_1 = 1, x'_2 = x_1, x'_3 = x_2, \dots$$

$$b'_{i1} = a_i, b'_{i2} = b_{i1}, b'_{i3} = b_{i2}, \dots$$

とする。

式2-22 を行列とベクトルを用いて書き直すと，

$$z = b' x' \quad 2-23$$

となる。すなわち，入力データに定数1を加えることにより，式2-21は式 2-23 と同等になる。MR型ネットワークは特別な場合（動作関数が線形関数の場合）に重回帰分析法と同じ近似レベルになる。また入力データに定数1を加えることにより第2層のニューロンの動作関数（式1-1）の  $\theta$  は  $w$  の要素に置き換えられるため，最適値は学習の過程で自動的に求まる。<sup>19)</sup>

## 2. 4 パーセプトロン型ネットワークの層数

パーセプトロン型ネットワークの分類及び fitting 能力はネットワークの中間層の数，それらの層のニューロン数，及びニューロンの動作関数に依存する。

入力層と出力層の2層のみからなるネットワークの分類能力には限界がある。

例えばexclusive ORのような線形分離不可能なパターン群の学習ができないことが知られている。線形分離不可能なパターンを学習するためには，ネットワーク内での入力情報の内部表現が必要である。そのためには中間層が1以上必要で

ある。中間層の数はニューロン数が無限に取れば1で良いが、有限個ならば2必要なこともある。ニューロン層数の増大はCPU時間を著しく増大させるので、構造活性相関の分野で実用的に使えるネットワーク構造を決定する意義は存在する。

本論文では、まず3層のネットワークを用い、構造活性相関への適用事例を検討する。ニューラルネットワークの構造は  $N(l, m, n)$  で表記する。ここで、 $l, m, n$  はそれぞれ第1層，第2層，および第3層のニューロン数である。

### 第3章 パーセプトロン型ネットワークの分類および fitting 特性

本章では薬物の構造活性相関への応用を念頭においたニューラルネットワークの分類および fitting 特性を述べる。ニューラルネットワークの基本的な動作は非線形である。線形動作と非線形動作の特徴を明確にし、非線形動作に伴う問題点を指摘する。

#### 3. 1 分類

ニューラルネットワークの学習の過程で、入出力データの関係に関する情報はニューロン間の結合荷重行列に蓄えられ、データはそれによって分類される。この機能は未学習データでも働く。これをニューラルネットの推論機能と言い、構造活性相関の研究に適用可能である。本節ではその根拠を議論する。

##### 3. 1. 1 バックプロパゲーションアルゴリズムの実施方法

バックプロパゲーションアルゴリズム (式2-9,10,11) を適用する際、決定しなければならぬ幾つかの要素がある。それはパラメータ  $\alpha$ ,  $\theta$ ,  $\epsilon$  値の選定, およびニューラルネットワークの構造である。また式2-3の誤差関数  $E$  の値を個々の入力データについて、どの程度まで小さくすればネットワークの分類機能が実現されるかを検討し、学習の完了条件を決定する必要がある。

ニューロンのパラメータ  $\alpha$  と  $w_{ij}$  値は式1-1,1-4より、結合してニューロンの出力値を与える。従って  $\alpha$  の値によってネットワークの分類結果が左右されることは無い。ただしCPU速度が十分でない計算機の場合、 $\alpha$  を0.7~5の範囲で可変にとって効率的な学習を行うことが可能である。これ以外の領域では数値計算精度に関する問題が発生し易い。

パラメータ  $\theta$  は第 1 層の一つ以上のニューロンを 0 でない定数値 (通常 1) を出力するものとする、第 2 層の各ニューロンについて  $\theta = 0$  とすることが出来る。この定数出力値を保持しているニューロンのことをバイアスと呼ぶ事がある。バイアスによって  $\theta = 0$  と出来ることは、式 1-1 により複数の第 1 層のニューロンの出力値が合成されて一つのニューロンに入力されることを考えれば明らかである。

2.3.2 節で MR ネットワークでは  $\theta$  が学習の過程で自動的に求まることを示したが、一般のパーセプトロン型ニューラルネットワークにおいてもそれは成立する。

第  $n$  層にバイアスを置けば第  $n + 1$  層のニューロンの  $\theta$  は 0 にできる。十分な数のニューロンを第 1 層以外の層に配置すれば、第 1 層以外にバイアスを導入する必要があることが経験的に知られている。これは必要以上のニューロンを入出力層以外に置くと 1 ないし複数のニューロンに結合している  $w$  の要素が 0 となり自動的にバイアスに相当するものが学習過程で生成するためである。この実例が第 5 章の再構築学習法の適用において見られる。

パラメータ  $\varepsilon$  は多次元空間をどれだけ極小解の方向に進ませるかを指示する係数である。 $\varepsilon$  はニューロンの動作関数にシグモイド関数を使用した場合は 0.1 ~ 0.2 に取られることが多いが、その数学的根拠は無い。小さな  $\varepsilon$  値を採用すれば極小解に到達する時間が大きくなり、大きい値を使用すれば極小解を飛び越し学習が収束せず振動する。

ニューラルネットワークの構造決定因子には、ネットワークの層数とニューロンの個数の 2 つの要素がある。ネットワークの層数は非線形分類を行うためには 3 層以上が必要である。普通パターン認識問題には 3 層が多く使用されているが、これは無限に多くの中間層を有するネットワークでは 3 層で任意の関数が近似出来るという証明<sup>40)</sup>があること、および 4 層以上では学習時間が長くなることに基づく。これらの根拠は積極的に 3 層構造を支持するものではないが、初めに 3 層で研究を試行するということは合理的である。薬物の構造活性相関に 3 層のニューラルネットワークを適用してみると、分類不可能な事例は発生せず、4 化合物、6 種の活性、173 データで 81% 以上の確率で



分類が正しく行われた。<sup>41)</sup> 100%の確率で分類出来ないのは、用いているデータそのものに矛盾を含むためである。データの矛盾については3.2.2節で議論する。このような場合に4層以上のネットワークの振る舞いを検討しなければならない必然性は少ないように思われる。本論文では3層のニューラルネットワークのみを検討の対象にする。

ニューロンの個数については第1, 3層についてはそれぞれ入力、教師データの種類数によって決まる。問題になるのは第2層のニューロンの個数であるが、現在のところその個数決定法について研究<sup>42)</sup> されてはいるが、まだ決定的なものは知られていない。判明しているのは次の点である。

1. 必要ニューロン数は入力、教師データの情報量に依存する。
2. 必要ニューロン数はバックプロパゲーション学習に付加条件を導入すると変化する。
3. ニューロン数を必要以上に多くしても、ネットワークの分類機能は大きく向上したり低下することが無い。

第2層のニューロンの個数を問題にするのは過学習の可能性があるためである。過学習とは、対象を表わす十分な学習データが利用出来ない場合、学習の進行に伴って学習データについての近似精度は向上していくが、未学習データに対する近似精度は悪化していく現象のことを言う。これはニューラルネットワークが学習の進行に伴って、学習データ固有の構造の細部を精密に近似してしまうため、対象が本来持っているマクロな構造の再現作用が劣化することが原因である。第2層のニューロン数が多い状況では過学習の恐れがあり、それを避けるために必要最少限のニューロン数を求める方法が求められてきた。

第2層のニューロン数は第5章の再構築学習法から求められる事もあるが、原始的ではあるが、初期値として第1層のニューロン数の2倍程度の値から、学習が出来なくなるまで減少させて探す方法が確実かつ速い。

以上述べたように、薬物の構造活性相関に使用出来るニューラルネットワークの

パラメータ，構造，収束条件を決定するために， $\epsilon$  値，第2層のニューロン数，E 値による結果の変動を，学習方程式が実用性を保つ範囲で，数値計算により調べることに意味があるが，その他の要素については調べる必要性は少ないと考える。

### 3. 1. 2 ニューラルネットワークによる分類

代表的な非線形分類問題として， $n$ 次元超立方体の各頂点に学習点を置き，その学習点の帰属クラスを2種類にし，その配置をexclusive OR型にする。これを $n$ 次元超立方体のexclusive OR学習と呼ぶ。次の漸化式，

$$L^1 = {}^t(0, 1),$$

$$L^{n+1} = \begin{pmatrix} L^n, 0 \\ L^n, 1 \end{pmatrix},$$

で定義される $2^n$ 行 $n$ 列の行列 $L^n$ において， $L^n$ の任意の1行を学習点の座標とし，その行の各要素 $L_i$ について， $\text{mod}(\sum L_i, 2)$ を帰属クラス値とする。0，1はそれぞれ0，1ベクトルである。ニューラルネットワークの入力は学習点の座標およびバイアスであり，教師データは帰属クラスとする。このとき式2-3のEが0.00025以下になるまでに要したバックプロパゲーションの計算回数を表3-1に示す。

表3-1において， $N$ は第2層のニューロン数である。XXXXは学習時間が極めて長時間になることを示し，.....は学習が振動し収束方向へ向かわなかったことを示す。

学習が収束するためには， $n=2, 3, 4$ のときそれぞれ， $N \geq 3, 6, 10$ が必要であった。

$n=2$ のとき $N \geq 3$ は妥当であるが， $n=3$ のとき $N \geq 4$ とならないのは，ある程度の広さの平面を区切るために2個のニューロンを必要とするためと思われる。 $n=4$ の時 $N \geq 10$ も同様である。経験的に知られているように，学習は $\epsilon$ が小さい領域では非効率になり，大きい領域では振動的になる。

表3-2に学習が完了したときのニューラルネットワークの出力値と教師データの差の2乗和を示す。

表3-1.  $n$ 次元超立方体のexclusive OR学習におけるバックプロパゲーション回数

$n = 2$

$\epsilon \setminus N$	3	4	5	6
0.05	135388	94223	108913	90823
0.10	86686	68555	49212	48845
0.15	57429	41042	47772	29254
0.20	36391	44421	30151	32180
0.25	43668	22145	24884	23323
0.30	.....	.....	29942	19940
0.35	.....	30801	22253	18572

$n = 3$

$\epsilon \setminus N$	6	7
0.10	37223	67158
0.15	83108	59728
0.20	54227	54366
0.25	40984	45738
0.30	33957	31660
0.35	41574	30879
0.40	.....	35563

$n = 4$

$\epsilon \setminus N$	10	11
0.15	102500	XXXX
0.20	88311	87212
0.25	.....	.....
0.30	.....	47487
0.35	53977	.....

表3-2.  $n$ 次元超立方体のexclusive OR学習時のニューラルネットワークの分類機能  
(期待値との差の2乗和)

$n = 2$

$\epsilon \setminus N$	3	4	5	6
0.05	0.00166	0.00169	0.00152	0.00132
0.10	0.00139	0.00115	0.00159	0.00186
0.15	0.00132	0.00137	0.00106	0.00156
0.20	0.00149	0.00124	0.00125	0.00142
0.25	0.00146	0.00136	0.00110	0.00113
0.30	....	....	0.00097	0.00109
0.35	....	0.00098	0.00103	0.00105

$n = 3$

$\epsilon \setminus N$	6	7
0.05	0.00196	0.00170
0.10	0.00135	0.00176
0.15	0.00162	0.00153
0.20	0.00161	0.00141
0.25	0.00114	0.00188
0.30	0.00116	0.00141
0.35	.....	0.00120

$n = 4$

$\epsilon \setminus N$	1 0	1 1
0.15	0.00153	XXXX
0.20	0.00140	0.00139
0.25	.....	.....
0.30	.....	0.00161
0.35	0.00130	.....

学習が完了すれば  $\epsilon$  , 第2層のニューロン数  $N$  にかかわらず, 同じ分類機能を持つ解に収斂する。期待値との差の2乗和は  $10^{-2}$  オーダであり, 十分な分類機能の実現されている。式2-3の  $E$  値に対するネットワークの分類機能の依存性は  $E \leq 0.15$  のとき存在しない。  $N$  による結果の依存性が殆ど見られないのは付録1に示すようなニューラルネットワークの剛性のためと考えられる。

### 3. 1. 3 ニューラルネットワークによる推測

ニューラルネットの推論機能を議論するためには, 多次元空間中の学習データ配置点の近傍についての知識が必要である。ニューラルネットワークの出力値を入力データ値について偏微分する。

式1-1を  $x_i$  について偏微分すると,  $\delta y_j = w_{ij} \delta x_i$ , 式1-4, 2-2から,  
 $\delta O_j = f' (y_j) \delta y_j$ .

これが各層で成立するので, 3層のネットワークでは,

$$\delta O_j / \delta x_i = \sum_k f' (y_k^{(2)}) w_{ik}^{(1,2)} f' (y_j^{(3)}) w_{kj}^{(2,3)}, \quad 3-1$$

である。ここで  $f'$  は  $f$  を  $y$  について微分した関数,  $f' (y_k^{(2)})$  は第2層の  $f'$  関数の  $y_k$  における値である。  $w_{ik}^{(1,2)}$  は1, 2層間のニューロン間結合荷重である。この式は,  $\delta O_j = \text{const.} \times \delta x_i$  であるから, 学習点を中心にした微小半径の超球領域内では学習点と同じ分類が行われることを示す。学習操作の差異によって, 同じ分類を行うネットワークでも各  $w_{ij}$  は互いに異なるので, 超球の半径は異なる。また学習点の位置によっても同じ分類が行われる超球領域の半径が異なる。これらは未学習点のニューラルネットワークの予測では, 学習点から離れた点の予測値の信頼性は低くなることを示す。

多次元空間における任意の2点と同じ分類に属する場合,  $w_{ik}^{(1,2)}$  が異なって同じ  $O_j$  を与えるためには,  $f' (y_k^{(2)}) = 0$ , または  $f' (y_j^{(3)}) = 0$  が成立するか,

あるいは偶然に式1-1, 1-2を組み合わせる3層のネットワークに適用した式が等価になるかのいずれかである。前者の条件が成立するとき、 $\delta O_j / \delta x_i = 0$ 、すなわち上記の2点を結ぶ線分上の未学習点に沿って、学習点と同じ分類が行われる領域が広がる。この条件は常に成立するわけではないが、成り立つ可能性は高く、構造活性相関の等級分類に有用な性質である。これは、構造活性相関に用いるニューラルネットワークのニューロンの動作関数には、 $f' = 0$ となる部分が含まれていることが望ましいことを示す。ニューラルネットワークの隠れ層のニューロン数の検討から、同じ結論が阿倍<sup>42)</sup>によって示されている。

### 3. 1. 4 未学習点の分類作用 (exclusive ORの場合)

ニューラルネットワークの未学習点での分類作用を数値計算によって調べる。計算例はn次元超立方体のexclusive ORの学習とする。exclusive OR系には次のような性質がある。ネットワークの入出力データ間の変換を $\Omega$ 、クラス1, 2に属する入力ベクトルを $x_1, x_2$ 、クラスの表現型を $t_1, t_2$ 、教師データベクトルの0, 1要素を交換する変換を $R$ とすると、 $\Omega x_1 = t_1, \Omega x_2 = t_2, R t_1 = t_2$ 、という関係が成立する。したがって、

$$\Omega x_2 = R (\Omega x_1) . \quad 3-2$$

この式は、クラス2に属する入力データの出力値はクラス1に属するデータの出力の0, 1を反転したものであることを示す。

$x_1$ から $\Delta_\nu$ だけ $x_2$ の方向に変位したベクトル $x_\nu$ を次のように定義する。

$$x_\nu = (1 - \Delta_\nu) x_1 + \Delta_\nu x_2, \quad \Delta_\nu = \{\delta_1, \delta_2, \dots, \delta_n\} .$$

このベクトルをニューラルネットワークに入力したときの $x_1$ 入力に対する出力の変化 $\xi$ は、式3-1より、演算子 $\Delta_\nu$ の1要素 $\delta_\mu$ について、

$$\xi_\mu = \sum f' (y_k^{(2)}) w_{ik}^{(1,2)} f' (y_j^{(3)}) w_{kj}^{(2,3)} \delta_\mu . \quad 3-3$$

式3-3において $w_{ik}^{(1,2)}$ 等は $x_1$ から $t_1$ を計算したときの結合荷重である。

同様に、 $x_2$ から $\Delta_\nu$ だけ $x_1$ の方向に変位したベクトル $\overline{x_\nu}$ をニューラルネットワークに  
入力したときの出力の変化 $\xi'$ は、

$$\xi'_\mu = \sum f'(\overline{y_k^{(2)}}) \overline{w_{ik}^{(1,2)}} f'(\overline{y_j^{(3)}}) \overline{w_{kj}^{(2,3)}} \delta_\mu, \quad 3-4$$

式3-3,4において、 $y_j^{(3)} = \overline{y_j^{(3)}}$ ,

$$\sum f'(\overline{y_k^{(2)}}) \overline{w_{ik}^{(1,2)}} \overline{w_{kj}^{(2,3)}} = \sum f'(\overline{y_k^{(2)}}) \overline{w_{ik}^{(1,2)}} \overline{w_{kj}^{(2,3)}},$$

よって $\xi = \xi'$ 。この考え方を0~1の区間に適用して、式3-1を、

$\Omega(\Pi \Delta_\nu) x_2 = R\{\Omega(\Pi \Delta_\nu) x_1\}$ 、と書くことが出来る。これはパス

$\Pi \Delta_\nu$ および $\Pi \Delta_\nu$ の対応する点 $\eta$ について、

$$\Omega \Delta_\eta x_2 = R(\Omega \Delta_\eta x_1) \quad 3-5$$

が成立することを示す。 $\Delta_\eta$ は $x_1, x_2$ から $x_\eta$ を生成する演算子である。式3-5は  
式3-2が成立するとき、学習の完了したニューラルネットワークの未学習データに対す  
る出力値の性質を記述している。

n次元超立方体のexclusive OR学習の、超立方体内部の点のニューラルネットワーク  
の出力値を表3-3Aに示す。点数は $n=2, 3, 4, 5$ に対しそれぞれ、 $100^2, 31^3, 20^4, 20^5$ で  
ある。この表は0.1単位に区分された区間にネットワークの出力値が入る割合を%で示  
したものである。表3-3Aでは学習が完了しなかった $\epsilon$ 値については省略されている。

$n$ は次元数、 $N$ は第2層のニューロン数を示す。式3-5により $t_2$ 側のニューロン値の分  
布は $t_1$ 側と反対称になるため省略した。 $u$ は区分された区間の上限値である。

ニューロン出力値の分布は $n, N, \epsilon$ の値によって変化する。またニューラルネット  
ワークのニューロン間の結合荷重の初期値によっても変化する。学習方程式は学習点間  
の出力値分布に関する制約を導入していないので、この変動は当然起こり得る。

exclusive OR系では同じクラスの学習点を結ぶ線分は必ず交わる。この線分のことを  
以下 $\phi$ 線分と呼ぶ。2つの $\phi$ 線分が交わっているとき、どちらかのクラスに属する $\phi$ 線  
分に沿って、片方のクラスに分類される領域が優越する。これを分類機能の非対称性と  
いう。 $n$ 次元超立方体のexclusive OR学習は、 $\phi$ 線分が必ず交わるようになっている

表3-3A. n次元超立方体のexclusive OR学習の場合の未学習点の  
ニューラルネットワーク出力値分布 (%)

n = 2

N = 3 :

$\varepsilon \backslash u$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	10.6	4.4	3.0	2.5	2.4	2.5	3.0	3.9	7.3	60.4
0.10	11.6	4.1	2.7	2.2	2.2	2.2	2.6	3.5	6.2	62.8
0.15	63.0	6.5	3.6	2.9	2.3	2.2	2.5	2.7	4.1	10.3
0.20	11.0	4.1	2.9	2.3	2.3	2.3	2.8	3.7	6.7	61.9
0.25	7.1	3.0	2.1	1.7	1.7	1.7	2.0	2.8	4.9	73.1

N = 4 :

0.05	8.5	4.2	2.8	2.5	2.6	2.2	3.2	4.2	7.6	62.0
0.10	11.5	5.3	3.7	3.2	3.2	3.4	4.0	5.4	10.3	49.9
0.15	52.9	9.4	5.3	3.8	3.3	3.0	3.3	3.6	5.2	10.2
0.20	63.6	6.4	3.7	2.5	2.2	2.2	2.5	2.8	3.8	10.5
0.25	14.0	6.0	4.4	3.9	3.9	4.4	5.6	8.9	21.8	27.3
0.35	14.4	5.0	3.4	2.9	2.8	2.9	3.5	4.7	9.0	51.6

N = 5 :

0.05	60.5	7.6	4.8	3.4	2.9	2.4	2.8	3.0	4.5	8.1
0.10	60.8	7.9	4.4	3.1	2.9	2.7	3.1	3.1	4.3	7.6
0.15	12.0	5.6	4.1	3.7	3.6	4.1	5.0	7.6	18.5	35.8
0.20	12.9	5.6	4.1	3.7	3.7	4.1	5.2	7.9	20.7	32.1
0.25	12.5	5.8	4.5	4.0	4.3	4.9	6.6	12.9	16.7	27.9
0.30	12.4	5.9	3.5	3.1	2.9	3.2	3.8	5.2	9.8	51.1
0.35	13.9	5.5	4.0	3.5	3.5	3.9	4.9	7.1	18.9	34.9

N = 6 :

0.05	24.1	20.2	10.2	6.5	5.2	4.7	4.6	5.1	6.8	12.5
0.10	9.0	4.3	3.0	2.5	2.4	2.6	3.0	4.1	7.5	61.6
0.15	26.7	21.5	9.4	6.1	5.0	4.4	4.4	4.8	6.4	11.4
0.20	60.9	7.4	4.3	3.2	2.8	2.6	2.7	3.1	4.4	8.8
0.25	13.9	6.6	5.1	4.8	5.1	6.5	11.1	10.7	12.9	23.4
0.30	14.2	6.7	5.2	4.8	5.1	6.3	10.5	10.5	12.6	24.2
0.35	14.3	5.9	4.4	3.9	4.1	4.6	5.9	9.8	20.3	26.8

u : 区間の最大値



n = 3

N = 6 :

$\epsilon \backslash u$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.10	17.0	12.3	10.3	11.5	9.2	7.0	6.3	6.9	9.2	10.2
0.15	34.6	3.4	3.2	3.2	0.6	2.7	3.2	2.8	4.5	41.9
0.20	34.4	6.2	2.7	0.8	3.2	3.1	0.6	3.3	6.6	39.4
0.25	34.4	5.5	2.4	1.8	3.0	2.0	2.0	3.1	6.0	39.8
0.30	32.3	5.6	3.1	1.7	1.8	3.2	3.2	3.0	6.6	39.3
0.35	34.0	4.0	2.7	0.9	3.2	0.5	2.8	3.3	4.0	44.6

N = 7 :

0.10	25.6	11.2	6.3	4.8	4.2	4.3	4.8	6.9	11.0	20.7
0.15	34.2	4.9	2.9	2.4	2.3	2.0	2.8	3.3	5.5	39.7
0.20	34.4	4.6	2.3	2.7	3.2	0.8	2.9	3.3	4.5	41.4
0.25	34.4	3.3	3.5	2.8	2.8	0.8	3.2	3.3	4.0	41.9
0.30	34.2	6.1	3.4	0.6	2.9	3.2	2.5	1.4	6.4	39.4
0.35	34.9	3.3	3.1	3.1	2.4	1.5	2.9	3.2	4.1	41.6
0.40	22.2	8.8	7.1	6.5	6.4	5.7	5.8	6.1	7.8	23.7

n = 4

N = 10 :

0.15	33.7	5.1	0.5	4.8	0.1	4.1	1.4	1.1	4.9	44.3
0.20	36.1	2.7	5.3	0.0	5.5	0.0	0.1	5.7	3.0	41.6
0.35	36.3	2.5	5.3	0.0	5.5	0.0	0.2	5.5	3.0	41.7

N = 11 :

0.20	33.8	5.0	5.3	0.0	0.7	4.8	0.2	5.6	0.1	44.4
0.30	31.7	4.9	2.6	2.7	2.5	1.0	2.9	1.7	5.7	44.4

n = 5

N = 14 :

0.20	41.5	4.1	3.6	0.9	1.7	2.9	0.8	3.2	3.0	38.3
------	------	-----	-----	-----	-----	-----	-----	-----	-----	------

u : 区間の最大値

単位 : %

特異な系であって、分類機能の非対称性が現れ易い。情報科学の分野でexclusive ORの学習が代表として選ばれているのは、線形分類不可能、かつこれ以上分類が困難な系はないためと思われる。

表3-3Aより次のような傾向が見られる。

1.  $n$ が大きくなる程分布が対称形になる。
2.  $n$ が大きくなる程未学習点の出力値が0, 1に近づく。
3.  $N$ が大きくなる程分布が対称形になる確率が高い。

表3-3Aから判断する限り、 $n \geq 4$ 、 $N \geq 10$ では問題なくニューラルネットワークで推測が可能な領域であるといえる。

### 3. 1. 5. 未学習点の分類作用 (4分類系の場合)

exclusive OR系以外の分類において、 $R t_2 \neq t_1$ 、 $R t_1 \neq t_2$ となる場合について検討する。解析を容易にするため、多くの分類問題の中から入力データ番号に関し対称性を有し、部分的にexclusive ORの性質を持っている系を選ぶことにする。その条件を満足する分類問題の一つとして、次の様に定義される入力、教師データ、

$$x_i = \{ (c_1, c_2, c_3, c_4), Q(c_1, c_2, c_3, c_4) \}, i = 1 \sim 8,$$

$$t_k = \{ \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14} \}, k = 1 \sim 4,$$

で定義される系を選んだ。ベクトル $c_k$ は $k$ というクラスに属する学習点の座標をあらわす。 $c_k$ の $z$ 成分は0である。 $\delta_{ik}$ はデルタ関数である。 $Q$ は $\{c_k\}$ を $\{c_{k'}\}$ とする変換を表わす。ここでは、 $c_1, c_2, c_3, c_4$ の $x, y$ 成分を、

$(0, 0), (1, 0), (1, 1), (0, 1)$ とする。 $Q$ の具体的表現は、

$$Q_1 = (1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 1, 4 \rightarrow 2; z_{+1}),$$

$$Q_2 = (1 \rightarrow 4, 2 \rightarrow 3, 3 \rightarrow 2, 4 \rightarrow 1; z_{+1}),$$

$$Q_3 = (1 \rightarrow 3, 2 \rightarrow 2, 3 \rightarrow 1, 4 \rightarrow 4; z_{+1}),$$

$$Q_4 = (1 \rightarrow 4, 2 \rightarrow 1, 3 \rightarrow 2, 4 \rightarrow 3; z_{+1}),$$

である。これら以外の $Q_n$ の定義もあるが、それらは $Q_1 \sim Q_4$ のどれかに帰属でき独立ではない。ここで $z_{+1}$ は $c_k$ 座標の $z$ 成分を $+1$ とすることを表わす。これを4分類系と呼ぶことにする。上記の例は3次元の場合である。より高次の場合を付録2に示す。

$x_1$ の中に含まれる $Q_n$ 変換の種類によって4分類系を識別する。

各分類の特徴は次の通りである。

$Q_1$ : 4つの $\phi$ 線分が一点で交わる。

$Q_2$ : 2つの $\phi$ 線分が一点で交わり、それが2組ある。

$Q_3$ : 2つの $\phi$ 線分が一点で交わる。

$Q_4$ : 4つの $\phi$ 線分がどれも交わらない。

$Q_2, Q_3$ はexclusive OR系と同じように二つの $\phi$ 線分が交わる場合である。

表3-3Bに、4分類系の学習における立方体内部の点のニューラルネットワークの値を0.1単位に区分された区間にネットワークの出力値が入る割合を%で示す。 $N=6$ ,  $\epsilon=0.25$ , 内部点数は $31^3$ である。 $u$ は区分された区間の上限値である。1つの $Q_n$ の中では $k$ の順序によって分布が示されている。

この表から次のことが観測される,

$Q_1$ : クラス2に属する $\phi$ 線分に沿った領域が他のクラスの領域より大きい。

$N, \epsilon$ を6~7,  $0.1 \leq \epsilon \leq 0.4$ の範囲で変化させ11種の相異なる組み合わせで計算したところ、クラス1, 2, 3, 4に属する領域がそれぞれ2, 2, 4, 3回優越となった。一つのクラスに属する領域が競合により優越となるが、クラス番号は学習条件により不定になっている。これは上記の $Q_1$ の特徴と矛盾しない。

$Q_2$ : クラス3, 4に属する2つの $\phi$ 線分に沿った領域が他の2クラスの領域より大きい。 $N, \epsilon$ を変化させて14種の組み合わせで計算すると、クラス対(3, 4), (1, 2), (1, 3), (2, 3)がそれぞれ7, 5, 1, 1回優越となった。 $Q_2$ の特徴は $\phi$ 線分の交わりが独立に2個存在することであるから、対番号不定で2つのクラスの領域が拡大する理由が理解出来る。

表3-3B. 4分類問題の場合の未学習点の  
ニューラルネットワーク出力値分布 (%)

$N = 6, \varepsilon = 0.25,$

$\backslash u$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$Q_1$	85.2	3.4	1.9	1.3	1.1	1.0	0.9	1.0	1.3	2.9
	14.8	5.8	4.2	3.6	3.5	3.7	4.2	5.4	8.8	46.0
	73.2	5.5	3.0	2.2	1.9	1.7	1.8	2.0	2.7	5.8
	84.2	3.6	2.0	1.4	1.2	1.1	1.0	1.2	1.5	2.9
$Q_2$	90.0	2.2	1.3	1.0	0.8	0.7	0.7	0.8	1.0	1.4
	79.2	5.0	2.7	1.9	1.6	1.5	1.5	1.6	1.9	2.8
	62.1	4.8	2.9	2.3	2.1	2.2	2.0	2.8	4.3	14.5
	54.8	7.1	4.4	3.5	3.0	3.0	3.2	3.5	4.9	12.5
$Q_3$	78.2	5.5	2.9	2.1	1.7	1.5	1.6	1.6	2.1	2.7
	72.7	4.2	2.4	1.9	1.5	1.5	1.6	1.9	2.8	9.5
	17.4	8.3	5.9	4.8	4.6	4.5	5.1	6.2	9.8	33.5
	73.6	4.3	2.4	1.9	1.6	1.5	1.6	1.7	2.6	8.7
$Q_4$	91.2	1.9	1.0	0.8	0.7	0.6	0.6	0.7	1.0	1.5
	51.1	8.1	5.3	4.2	3.6	3.4	3.4	3.6	4.8	12.4
	33.6	7.0	5.0	4.3	4.1	4.2	4.5	5.3	7.6	24.3
	81.5	4.0	2.2	1.6	1.4	1.2	1.3	1.4	1.9	3.6

$N$  : 第2層のニューロンの個数

$u$  : 区間の最大値

$Q_n$  : 置換の種類, 一つの置換の中ではクラス番号順に分布 (%) が表示されている。

Q<sub>3</sub>: クラス3に属する $\phi$ 線分に沿った領域が大きく、クラス1に属する領域が小さくなっている。クラス1と3の $\phi$ 線分は交わるので、これは予想出来る。N,  $\epsilon$ を変化させて14種の組み合わせで計算した場合、このタイプの分布が11回、全体に平均化した場合が3回であった。優越となるクラスの番号は1, 3が6, 5回ずつであった。これらの結果はQ<sub>3</sub>の特徴と矛盾しない。

Q<sub>4</sub>: クラス2, 3に属する2つの $\phi$ 線分に沿った領域が他の2クラスの領域より大きい。Q<sub>2</sub>の特徴に類似しているが、 $\phi$ 線分は交わってはいない。N,  $\epsilon$ を変化させた24種の組み合わせ計算の結果、優越クラス対は(2, 3), (3, 4)がそれぞれ23, 1回であった。この対番号の再現確率の高さはQ<sub>2</sub>とは異質のものである。従って複数のクラスの競合によって分布の偏りが生じたのではなく、他の原因を考える必要がある。

4分類系の場合、入力データを他のデータと交換する変換Tについて、

$\Omega(Tx_i) = \Omega x_i$ , が成立しなければならないが、表3-3Bからこの関係は明らかに成立しない。この現象を説明するには、学習の効率という考えが必要である(3.3.1節)。

4分類系においてもニューロン出力値の分布には大きな偏りがある。 $\phi$ 線分上の推測点以外の予想については、学習条件による多くの不確定要素が存在する。従ってニューラルネットワークで推測可能な領域は、 $\phi$ 線分上の点の他は、学習点の近傍の小さな領域であるといえる。

### 3. 2 F i t t i n g

Q SAR解析に適したニューラルネットワークは、2.3.2節で示したMR型ネットワークである。ここではQ SARに頻繁に用いられている $y = x$ ,  $y = |x - a|$ 関数からサンプリングしたデータを用いて、MR型ネットワークの動作特性を調べる。

### 3. 2. 1 直線 ( $y = x$ ) における fitting

MR型ネットワークへの入力として $y = x$ を、教師データとして $y = x$ をそれぞれ21点で離散化したデータを用いた。前者は測定の時間間隔に、後者は測定値に対応する。

MRネットワークの $\epsilon$ 値は第1, 2層間と2, 3層間で異なる値を用いた方が学習速度の面から望ましいことが経験的に知られている。 $\epsilon^{(2)} \sim 2 \times \epsilon^{(3)}$ にとって不都合はなかった。以下の研究ではそのようにとり、第3層の $\epsilon^{(3)}$ を $\epsilon$ とすることにする。 $\epsilon$ および $N$ を変化させ、 $E < 0.0001$ が満足されるまでに要したバックプロパゲーションの計算回数を表3-4に示す。XXXXは学習速度が遅いため計算を途中で中止させたことを示す。学習完了後のニューラルネットワークの出力値と教師データとの差の2乗和を表3-5に示す。

学習の完了したネットワークでは目的とした関数機能が $\epsilon$ 、 $N$ の値に依存せず実現されている。 $\epsilon$ 値を0.15以上にすると学習速度は速くなるが、生成するネットワークの機能は不十分である。 $y = x$ のような単純な関数は $\epsilon$ を大きくとることにより学習速度は速くなるが、Q S A Rに適用するには小さめの値( $\sim 0.05$ )を設定するのが安全ではないと思われる。

MR型ネットワークの内挿機能を次の様にして調べた。0.1区間に211点を取り各点でのネットワークの出力値を計算し、 $y = x$ との差を求めた。この差の分布を表3-6に示す。区間は $(\min, \max \text{値}) \times 10^{-3}$ で示し、 $k$ はその区間に入った差の数である。

差の分布は正規分布とは異なっている。ニューラルネットワークの内挿精度は優れている。この精度は目的とする関数が $y = x$ であつても保たれていた。

次にMR型ニューラルネットワークの外挿機能を調べた。 $\epsilon = 0.05$ 、 $N = 2$ のとき、入力データ値1.24で $-5\%$ 、1.43で $-10\%$ の誤差を生じ、 $N = 5$ では、1.29で $-5\%$ 、1.48で $-10\%$ の誤差であつた。外挿の場合は内挿に比べると精度は劣るが、それでも最大の測定範囲 $+20\%$ までは注意して使用できる範囲と思われる。

表3-4. MR型ニューラルネットワークの学習回数 ( $y = x$ )

$\varepsilon \setminus N$	2	3	4	5
0.005	XXXX	5441	6673	9999
0.01	XXXX	17033	3675	7147
0.02	37300	2810	2021	2125
0.05	11888	244	756	690
0.10	10310	163	110	350
0.15	9602	5303	7026	191
0.20	6158	6317	127	108
0.25	4824	53	51	69

表3-5. MR型ニューラルネットワークの出力値と期待値の差の2乗和 ( $y = x$ )

$\varepsilon \setminus N$	2	3	4	5
0.005	XXXX	0.00057	0.00056	0.00063
0.01	XXXX	0.00088	0.00055	0.00059
0.02	0.00112	0.00055	0.00057	0.00058
0.05	0.00079	0.00074	0.00060	0.00063
0.10	0.00069	0.00075	0.00090	0.00088
0.15	0.00064	0.00065	0.00061	0.00146
0.20	0.00063	0.00060	0.00195	0.00243
0.25	0.00061	0.00168	0.00372	0.00349

表3-6. MR型ニューラルネットワークの出力値と期待値の差の分布 ( $y = x$ )

min	max	k	min	max	k
-15	-9	7	0	1	9
-9	-8	2	1	2	9
-8	-7	2	2	3	9
-7	-6	26	3	4	10
-6	-5	17	4	5	11
-5	-4	12	5	6	12
-4	-3	10	6	7	16
-3	-2	10	7	8	27
-2	-1	9	8	9	1
-1	0	10	9	10	2

min: 区間の最小値, 単位は  $10^{-3}$ ,

max: 区間の最大値, 単位は  $10^{-3}$

k: (min, max) 区間に入る差の数

最大の出力値を $O_{max}$ とし、これを与える入力データを $x_i$ とする。 $x_i$ の正方向の微小変化を $\Delta$ とする。 $x_i \rightarrow x_i + \Delta \equiv x_{ex}$ の時の式2-2は、

$$O_{ex} = f(y_{ex}),$$

$$y_{ex} = \sum w_{i,max} x_{ex} = \sum w_{i,max} (x_i + \Delta) = y_{max} + (\sum w_{i,max}) \Delta, \quad 3-6$$

となる。従って $ex$ 点での偏微分係数は、

$$\begin{aligned} \delta O_{ex} / \delta x_{ex} &= f'(y_{ex}) w_{i,max} \\ &= f'(y_{max}) w_{i,max} + f''(y_{max}) w_{i,max} (\sum w_{i,max}) \Delta, \end{aligned} \quad 3-7$$

である。 $f'(y_{max}) w_{i,max} = f'(y_{max}) = 1$ に注意して、次の関係を得る。

$$\begin{aligned} \delta O_{ex} / \delta x_{ex} &= f'(y_{max}) w_{i,max} \{1 + (\sum w_{i,max}) \Delta \\ &\quad - 2 f'(y_{max}) (\sum w_{i,max}) \Delta\} \sim 1 - (\sum w_{i,max}) \Delta. \end{aligned} \quad 3-8$$

ここで、 $\Delta$ は小さい正数、 $\sum w_{i,max}$ は正の数であるので、

$$\delta O_{max} / \delta x_i > \delta O_{ex} / \delta x_{ex} > 0, \quad 3-9$$

という関係が $\Delta$ の小さい間は成り立つ。従って外挿においてもネットワークの出力値は単調増加である。この性質によってMR型ニューラルネットワークを用いてより強い生物活性を有する未知化合物の予測が可能である。

### 3. 2. 2 $y = k |x - a|$ におけるfitting

$y = k |x - a|$  という関数は生物活性と構造データ間にしばしば現れる関係である。 $k$ は定数である。この関数は $x = a$ で微分値が一義的でない。その点ではニューラルネットワークの学習方向が定義できず、ネットワークのfitting機能が十分ではなくなる。しかしどの程度不十分なのか数値計算で調べる必要がある。

MR型ニューラルネットワークの入力データとして $y = x$ を、教師データとして $y = |x - 0.5|$ を用い(0,1)区間を21点でサンプリングした。学習は $\epsilon$ ,  $N$ を変化させ、 $E < 0.005$ まで行った。結果を表3-7および3-8に示す。連続微分可能な関数の



場合と異なり、 $E$ をその値以上小さくするためには大きな $N$ 値と多くのCPU時間を要する。

学習の完了したネットワークでは $\varepsilon$ および $N$ の値に依存せず目的とした関数機能が実現されている。 $\varepsilon$ を0.20以上にすると学習速度は速いが、学習が振動し完了しないこともある。0.02以下では学習速度が極めて遅く実用にならない。

学習完了時のネットワークの内挿機能を次の様にして調べた。0,1区間に211点を取りその各点のネットワークの出力値を計算し、 $y = |x - 0.5|$ との差を求めた。この差の分布は表3-9である。区間は $(\min, \max \text{値}) * 10^{-2}$ で示されている。

$y = |x - 0.5|$ の場合、 $y = x$ のような全区間学習可能な関数と異なり、MR型ニューラルネットワークのfitting能力は1オーダー程度低下する。真の関数値との差は正規分布ではない。

この型の関数を構造活性相関研究に使用する目的は、最大および最小値を与える入力データを求めることにあると思われる。ニューラルネットワークの出力から、それらを計算し表3-10に示した。関数域0~1の範囲において、 $10^{-3}$ の精度で最大および最小値を与える入力データが求めることが出来、MR型ニューラルネットワークの構造活性相関研究への適用には問題がないのではないかとと思われる。

### 3. 3 ニューラルネットワーク法の問題点

#### 3. 3. 1 非対称性分類の問題

ニューラルネットワークの分類においてみられる非対称性の問題について検討する。

非対称性は2つの要因に由来している。一つは、学習方程式は学習点以外については規定がないため、 $\phi$ 線分の交わりによる学習点間の分類機能が複数種実現される可能性である。これはいわゆるローカルミニマムに収束する問題とは異なる。

表3-7. MR型ニューラルネットワークの学習回数 ( $y=|x-0.5|$ )

$\epsilon \setminus N$	6	7	8	9
0.01	XXXX	45361	XXXX	XXXX
0.02	23383	42789	XXXX	26313
0.05	5212	14137	10705	23768
0.10	19085	6707	6981	18600
0.15	4342	2507	2012	13719
0.20	....	3780	1114	4168
0.25	....	2397	5773	3230

表3-8. MR型ニューラルネットワークの出力値と期待値の差の2乗和 ( $y=|x-0.5|$ )

$\epsilon \setminus N$	6	7	8	9
0.01	XXXX	0.0204	XXXX	XXXX
0.02	0.0198	0.0206	XXXX	0.0198
0.05	0.0254	0.0199	0.0185	0.0279
0.10	0.0197	0.0184	0.0183	0.0189
0.15	0.0326	0.0391	0.0404	0.0191
0.20	....	0.0184	0.0376	0.0330
0.25	....	0.0190	0.0221	0.0192

表3-9. MR型ニューラルネットワークの出力値と期待値の差の分布 ( $y=|x-0.5|$ )

min	max	k	min	max	k
-10	-9	0	0	1	23
-9	-8	0	1	2	30
-8	-7	0	2	3	35
-7	-6	0	3	4	7
-6	-5	0	4	5	3
-5	-4	6	5	6	2
-4	-3	28	6	7	2
-3	-2	25	7	8	2
-2	-1	24	8	9	3
-1	0	21	9	10	0

min: 区間の最小値, 単位は  $10^{-2}$ ,

max: 区間の最大値, 単位は  $10^{-2}$

k: (min, max) 区間に入る差の数

表3-10.  $y=|x-0.5|$ の最小値を与える  $x$  と  
ニューラルネットワークから計算された  $x$  値の差

$\epsilon \setminus N$	6	7	8	9
0.01	XXXX	-0.0005	XXXX	XXXX
0.02	-0.0005	-0.0010	XXXX	-0.0005
0.05	-0.0014	-0.0005	-0.0005	-0.0005
0.10	-0.0005	0.0000	0.0000	-0.0005
0.15	-0.0005	-0.0010	-0.0014	-0.0005
0.20	.....	0.0000	-0.0014	0.0000
0.25	.....	0.0000	0.0005	0.0000

N : 第2層のニューロン数

もう一つは式2-9に見られるデルタルールの非対称性である。すなわち、

$x_1, x_2$ という入力データがクラス $t_1, t_2$ に属し、 $t_1 \neq t_2$ かつ $x_1 < x_2$ のとき、

$x_1, x_2$ の中間点 $x_i$ を入力したときニューラルネットワークが $t_1, t_2$ の中間の値 $O_i$

を出力したとする。このとき、 $d_1 = (O_i - t_1) f'(y_i)$ 、

$d_2 = (O_i - t_2) f'(y_i)$ 、ゆえに $d_1 = d_2$ 。したがって式2-9より、

$\delta w_{ij}(at x_1) < \delta w_{ij}(at x_2)$ である。ここで $x_i$ 入力時を $at x_1$ と書く。

次のサイクルでの各変数は<sup>(2)</sup>で識別する。

$$\begin{aligned} \delta y_i^{(2)}(at x_1) &= \delta w_{ij}(at x_1) x_1 = -d_1 x_1^2 \varepsilon \\ &= -(O_i - t_1) f'(y_i) x_1^2 \varepsilon, \end{aligned} \quad 3-10$$

$$\begin{aligned} f'(y_i^{(2)}) &= f'(y_i) + f''(y_i) \delta y_i^{(2)}(at x_1) \\ &= f'(y_i) [1 - f''(y_i) (O_i - t_1) x_1^2 \varepsilon], \end{aligned} \quad 3-11$$

$$\begin{aligned} \delta O_i^{(2)}(at x_1) &= f'(y_i^{(2)}) \delta y_i^{(2)}(at x_1) \\ &\sim [f'(y_i)]^2 (O_i - t_1) x_1^2 \varepsilon, \end{aligned} \quad 3-12$$

$$\begin{aligned} d_1^{(2)} &= (O_i^{(2)} - t_1) f'(y_i^{(2)}) \\ &= \{O_i - t_1 + \delta O_i^{(2)}(at x_1)\} f'(y_i^{(2)}) \\ &\sim (O_i - t_1) f'(y_i) \{1 \\ &\quad - x_1^2 \varepsilon [(O_i - t_1) f''(y_i) + [f'(y_i)]^2]\} \end{aligned} \quad 3-13$$

[] 内の値が $\sim 1$ のような学習の初期では、 $\delta w_{ij}^{(2)}(at x_1) > \delta w_{ij}^{(2)}(at x_2)$

であるが、 $\varepsilon$ が小さくなると [] 内が $\sim 0$ となり、 $\delta w_{ij}^{(2)}(at x_1) <$

$\delta w_{ij}^{(2)}(at x_2)$ となる。すなわち入力ベクトルの大きな値を持つクラスの方が

効率的な学習が行われる。

この非対称性を除去することは3.1.4節に示したような問題に対称性がないと困難であるが、軽微にすることは可能である。以下に述べる方法はその一例である。

入力パターン(ベクトル)  $\{p_i\}$ 、及び教師パターン  $\{t_i\}$ 、の各要素は0と1の間にスケールされているものとする。このとき  $\{p_i\}$  および  $\{t_i\}$  の逆ベクトル

$\{\overline{p_i}\} = \{1 - p_i\}$ ,  $\{\overline{t_i}\} = \{1 - t_i\}$  を定義する。

これらのベクトルを用いると、次の4組のバックプロパゲーション学習の組み合わせが考えられる。

$$\begin{aligned} 1: & [\{x\}, \{t\}], & 2: & [\{x\}, \{\overline{t}\}], \\ 3: & [\{x\}, \{\overline{t}\}], & 4: & [\{\overline{x}\}, \{\overline{t}\}] \end{aligned}$$

これらのうち入力データの性質によっては等価な学習がある。等価な学習を除き有効な学習の数をNとする。N個のバックプロパゲーションの結果得られたネットワークの出力を次の式で合成する。

$$O_j' = N^{-1} \left\{ \sum_{I \in 1,2} O_{Ij} + \sum_{I \in 3,4} (1 - O_{Ij}) \right\} \quad 3-14$$

ここでIはバックプロパゲーションの組の番号、 $O_{Ij}$ はその組に属するニューロンの出力値を表わす。この一連の操作を対称学習<sup>20)</sup>と命名する。

もし入力、教師データ間にexclusive ORのような対称性があるのならば、3.1.4節の表記で  $(\Omega x_1 + \Omega x_2) / 2$ 、とすればパス  $\Pi \Delta_\nu$ ,  $\Pi \Delta_\nu$  の集合で定義される超平面について対称な分類機能が実現できる。

### 3. 3. 2 SARデータの特徴

ここでSARデータの特徴について議論する。SARの教師データは生理活性の強度順に、 $t_1 < t_2 < \dots < t_n$ 、の様に順序づけられたデータである。入力データはiという物理化学的な性質について、 $x_{\mu j} = f_\mu(t_j)$ 、なる関数関係を満足する。このとき2つの $\mu$ ,  $\nu$ に関する $\phi$ 線分が交わる可能性は、 $f_\mu = f_\nu$ あるいは $f_\mu = -f_\nu$ でありかつ、 $\mu$ ,  $\nu$ に対して別々のスケーリングを行った時である。この条件が共に成立することは無い。従ってSARにニューラルネットワークを適用した場合、 $\phi$ 線分交叉による推測機能の不確定性は考慮しなくて良い。

3.1.3節の議論から、等級 $t_j$ に帰属すべき入力データ $\mu$ の範囲が $(x_{\mu, \min}, x_{\mu, \max})$ であるとき、 $x_{\mu, j}$ が $x_{\mu, \min} \leq x_{\mu, j} \leq x_{\mu, \max}$ ならば、 $\phi$ 線分に沿って $t_j$ の領域が存在するのでほぼ正確な推測が可能である。 $x_{\mu, j}$ が $x_{\mu, j} < x_{\mu, \min}$ ,  $x_{\mu, \max} < x_{\mu, j}$ ならば、点 $x_{\mu, \min}$ ,  $x_{\mu, \max}$ のごく近傍では正しい推測が出来るが、それ以外では3.3.1節の議論から、入力ベクトルの大きな値の等級側に片寄った推測になる。

付録3に示した微分バックプロパゲーション法を適用すると、或る特定の等級間について片寄りの是正は可能である。但しその適用を行わなくともSAR教師データの性質により、最大誤差は+1等級である。

### 3.3.3 収束の困難な学習の問題

学習は設定され誤差関数値の閾値以内に至るまで繰り返されるが、極端に小さい閾値を設定すると学習は完了しない。パーセプトロンの収束定理により、無限のニューロンを用いれば学習は理論的には完了するが、シミュレータでは有限のニューロン数、計算時間、数値計算誤差の存在下で学習するので、閾値の最小値が存在する、それは経験的に $10^{-5} \sim 10^{-6}$ である。

ニューラルネットワークでの学習が困難な問題がある。これらの問題はしばしば薬物の構造活性相関に現れ、誤差関数 $E$ の微分値が一義的でなく学習方向が一定でなくなることが原因である。現象として、分類と関数fittingでは次の様に顕れる。

分類における学習不可能とは、入力データ $(x_i, x_j)$ 、教師データ $(t_1, t_2)$ に於いて、 $x_i = x_j$ ,  $t_1 \neq t_2$ となるデータを含む場合である。この場合、式2-3の $E$ は $\sim 0.5$ となり学習は完了しない。これをデータ関係の矛盾と呼ぶ。このとき、 $x_i, x_j$ を除くデータ関係の矛盾のない部分集合 $\{x\}$ ,  $\{t\}$ については学習が行われ、正しい分類が行われる。データ関係の矛盾が存在する場合、

all data

$$E_{av.} = N^{-1} \sum_i E_i$$

3-15

i

を学習完了の指標とすることが可能である。ここで、Nは全データ数である。

fitting問題において学習不可能な場合は、目的関数の微分値が一義に定まらない場合に見られる。このとき学習不可能点が関数の局所に限定されていれば、ニューラルネットワークによって実用に支障の無い精度のfittingが実現出来る。このとき学習に用いる閾値は最大 $10^{-2}$ 程度である。これ以上の閾値を用いなければならない問題は、MR型ニューラルネットワークで処理しても有用な結果は得られない。

### 3.3.4 因子解析

構造活性相関の研究では入力パラメータの薬物の生物活性の強度に及ぼす程度を知る必要がある。これは因子解析と呼ばれている。重回帰分析法のような線形解析では相関係数を解析することによってこの目的は達成できる。しかし、非線形動作をするニューラルネットワークでは相関係数の意味はなくなるため、これに代わる何らかの方法を考える必要がある。この問題は第5章で取り上げる。

### 3.3.5 ネットワーク出力の不変性

構造活性相関のデータには無視出来ない誤差が含まれている。この誤差を考慮したとき、パーセプトロン型ニューラルネットワークの出力値がどのように変化するかを考察する必要がある。この問題の詳細な議論は付録5に示すが、結論は、パーセプトロン型ニューラルネットワークの分類機能は重回帰分析法と比較して誤差の影響を受けにくいことが証明される。

### 3.4 leave-one-out法

分類および関数fittingにおける予測の妥当性の尺度として、leave-one-out法がしばしば用いられている。測定データがほぼ線形の関係にあるとき、それを1次多項式の識別関数で予想する場合、leave-one-out法は妥当な判定法の一つである。

ニューラルネットワークにおいてleave-one-out法を形式的に実施すると、次のようになる。

除去する学習点の入力データを $x_L$ 、教師データを $t_L$ とする。 $x_L$ が任意の他の学習点の入力データ $x_K$ および $x_M$ について $x_K < x_L < x_M$ となる状況にある場合、3.1.3および3.2.1節における議論により、

$t_K = t_L = t_M$ ならば、 $x_K, x_M$ 点を結ぶ $\phi$ 線分が $t_{K'} = t_{M'} \neq t_K$ を満足する $x_{K'}, x_{M'}$ 点を結ぶ $\phi$ 線分と交わらなければ、 $x_L$ を除去した学習結果から $t_L$ を予想出来る確率は高い。

これ以外の場合は $x_L$ を除去すると全く異なる学習を行うことになり、予想は当然出来ない。以上の理由により、ニューラルネットワークにおいてleave-one-out法を形式的に実施しても、ニューラルネットワークの予測機能を判定する手段にはなっていないことが判る。ニューラルネットワークを用いてleave-one-out法を重回帰分析と同じような意味で予想判定の手段とするためには、対象データが同一のカテゴリーに属する範囲でかつ $x_K < x_L < x_M$ を満たさない集合の外周に属する $x_L$ を除くleave-one-out操作を行う必要がある。



## 第4章 ニューラルネットワーク法の活性相関研究への応用

### 4.1 等級活性分類への応用

ニューラルネットワーク法の分類動作の応用として、ノルボルネン誘導体の $^{13}\text{C}$ -NMR ケミカルシフトと誘導体の2種のコンホメーション (endo, exo) の関係<sup>21)</sup>を学習し、未学習の誘導体のコンホメーションを予測する問題と、マイトマイシン誘導体<sup>22)</sup>及びアリルアクリロイルピペラジン誘導体<sup>23)</sup>の生物活性の等級分類及び予測問題を研究した。

#### 4.1.1 ノルボルネンの $^{13}\text{C}$ -NMR ケミカルシフトとコンホメーションとの関係

用いたノルボルネン誘導体を図4-1 に示す。表4-1 に誘導体のノルボルネン骨格の各炭素原子 ( $\text{C}_1$ - $\text{C}_7$ ) の $^{13}\text{C}$ -NMR ケミカルシフトのデータを示す。そこに示された数値は対応する誘導体の置換基のない基本骨格の化合物からのケミカルシフト差である。

ニューラルネットワーク構造は  $N(7,14,4)$  とし、第2層のニューロンの  $\alpha$ ,  $\epsilon$  及び  $\theta$  の値を 1.5, 0.2, 0.0 とし、第3層のそれらの値を 3.0, 0.2, 0.0 とした。(なお、これらの値を 1.0, 0.1, 0.0 としても結果は変わらないことを確認した。) 出力層のニューロン数を 4としたのは、出力への無意味な数値の混入の程度を調べるためである。教師パターンとして、endo には (0,1,1,0) を、また exo に対しては (1,0,0,1) を提示した。表4-2 は学習の繰り返し回数60での出力結果である。化合物番号1~25 は学習に用い、26~38 は未知データとして予測させた結果である。この程度の学習回数においても、出力ベクトルは (1,0,0,1) または (0,1,1,0) に十分近く、分類は完全に行われている。未知化

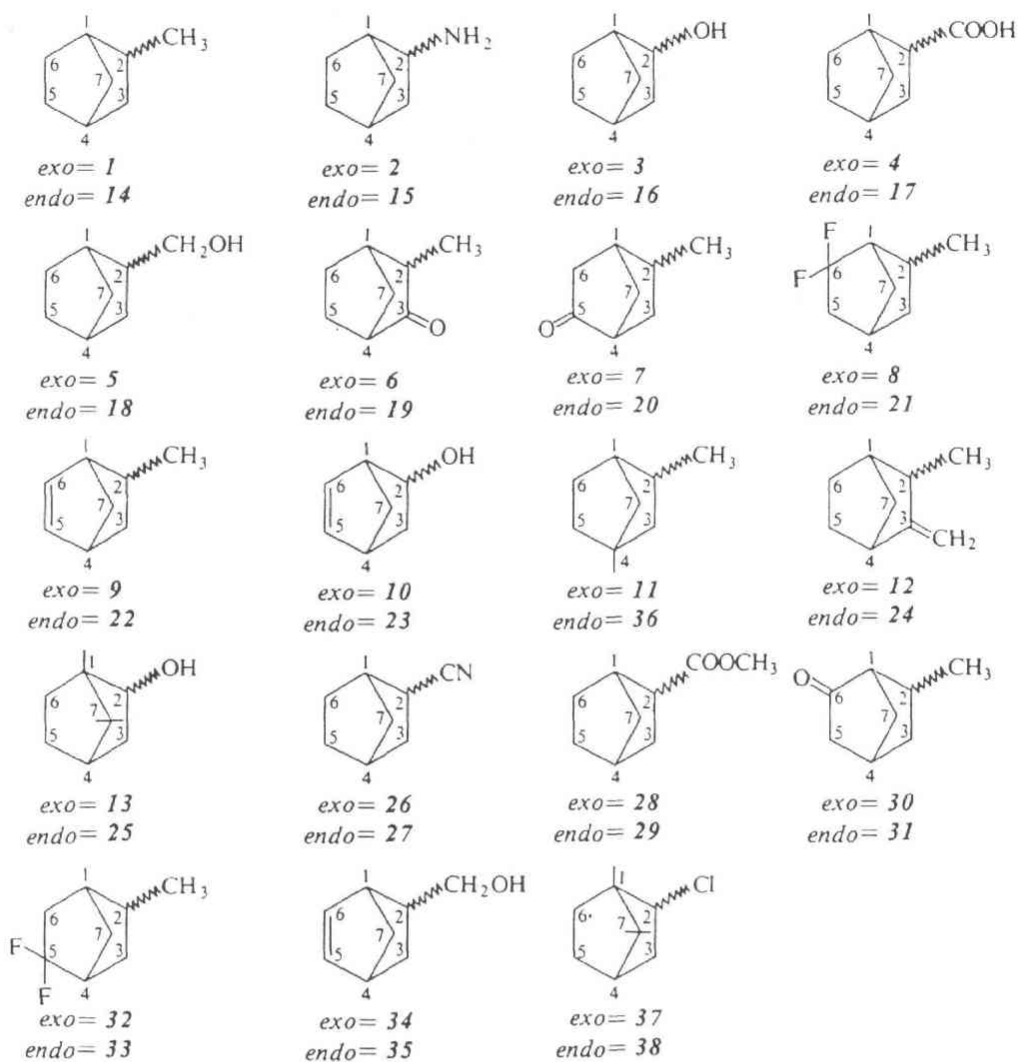


図4-1. パターン認識の対象としたノルボルネン類

表4-1. ノルボルネンの $^{13}\text{C}$ -NMRケミカルシフト値とコンホメーション

compd	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	exo/endo
1	6.7	6.7	10.1	0.5	0.2	-1.1	-3.7	exo
2	8.9	25.3	12.4	-0.4	-1.2	-3.1	-4.4	exo
3	7.7	44.3	12.3	-1.0	-1.3	-5.2	-4.4	exo
4	4.6	16.7	4.4	-0.2	-0.3	-1.0	-1.8	exo
5	1.8	15.1	4.4	-0.2	0.2	-0.7	-3.3	exo
6	5.7	3.0	2.6	-0.5	-0.4	0.7	-3.5	exo
7	6.1	5.9	10.6	0.6	0.2	0.2	-3.7	exo
8	6.5	6.3	10.4	0.3	-0.8	-0.1	-3.5	exo
9	6.5	7.5	9.5	0.5	1.7	0.7	-3.8	exo
10	7.8	47.0	11.7	-1.3	3.9	-2.7	-3.2	exo
11	6.9	6.9	10.1	0.7	-1.2	0.1	-3.9	exo
12	5.6	4.9	7.0	0.2	-1.1	0.2	-3.9	exo
13	2.5	42.5	11.9	-0.8	-1.1	-2.4	1.4	exo
14	5.4	4.5	10.6	1.4	0.5	-7.7	0.2	endo
15	6.8	23.3	10.5	1.2	0.6	-9.5	0.3	endo
16	6.3	42.4	9.5	0.9	0.2	-9.7	-0.9	endo
17	4.2	16.2	2.1	0.9	-0.6	-4.8	1.9	endo
18	1.7	12.8	4.0	0.4	0.2	-7.2	1.4	endo
19	4.7	3.1	2.2	0.3	1.3	-6.5	-0.6	endo
20	4.7	5.3	9.2	1.3	-0.4	-6.5	1.4	endo
21	4.6	11.5	8.9	-0.1	0.8	0.4	1.8	endo
22	5.6	7.5	8.7	1.4	1.7	-3.0	1.7	endo
23	7.1	47.8	13.3	2.2	3.6	-3.4	0.3	endo
24	4.1	4.2	7.0	0.7	0.5	-7.4	0.0	endo
25	3.2	40.2	10.4	-0.5	0.0	-10.3	3.1	endo
26	5.5	1.0	6.3	-0.3	-1.5	-1.6	-1.3	exo
27	3.4	0.1	5.5	0.2	-0.7	-4.9	0.0	endo
28	5.1	16.4	4.2	-0.4	-1.1	-1.4	-2.1	exo
29	4.0	15.9	2.2	0.7	-0.7	-5.0	1.7	endo
30	6.6	7.0	10.1	0.2	-1.2	0.5	-3.7	exo
31	6.0	8.4	11.2	-0.1	0.7	-1.5	-1.6	endo
32	6.3	7.2	9.8	0.7	-0.1	0.8	-3.5	exo
33	5.1	4.8	8.4	1.1	-0.1	-7.3	1.6	endo
34	1.9	17.1	5.2	-0.1	0.9	0.9	-3.4	exo
35	2.3	18.3	5.0	0.3	1.3	-2.9	1.4	endo
36	5.1	4.0	8.4	1.1	0.2	-7.7	1.6	endo
37	2.9	30.3	13.4	-0.5	-2.1	-0.7	2.0	exo
38	3.7	29.8	10.8	-1.6	-1.1	-9.0	2.2	endo

(点線以下は未学習データ)

表4-2. ニューラルネットワークの出力値とコンホメーション

compd	1	2	3	4	endo/exo
1	0.97	0.02	0.02	0.97	exo
2	1.00	0.00	0.00	1.00	exo
3	1.00	0.00	0.00	1.00	exo
4	0.94	0.06	0.06	0.94	exo
5	0.99	0.01	0.01	0.94	exo
6	0.99	0.01	0.01	0.99	exo
7	0.99	0.01	0.01	0.99	exo
8	0.99	0.01	0.01	0.99	exo
9	0.97	0.03	0.03	0.97	exo
10	0.99	0.01	0.01	0.99	exo
11	1.00	0.00	0.00	1.00	exo
12	1.00	0.00	0.00	1.00	exo
13	0.95	0.05	0.05	0.95	exo
14	0.00	1.00	1.00	0.00	endo
15	0.00	1.00	1.00	0.00	endo
16	0.03	0.96	0.96	0.03	endo
17	0.00	1.00	1.00	0.00	endo
18	0.00	1.00	1.00	0.00	endo
19	0.00	1.00	1.00	0.00	endo
20	0.00	1.00	1.00	0.00	endo
21	0.08	0.93	0.93	0.08	endo
22	0.00	1.00	1.00	0.00	endo
23	0.01	0.99	0.99	0.01	endo
24	0.00	1.00	1.00	0.00	endo
25	0.00	1.00	1.00	0.00	endo
26	0.83	0.18	0.18	0.83	exo
27	0.02	0.98	0.98	0.02	endo
28	0.98	0.02	0.02	0.98	exo
29	0.00	1.00	1.00	0.00	endo
30	1.00	0.00	0.00	1.00	exo
31	0.74	0.24	0.24	0.74	exo*
32	0.99	0.01	0.01	0.99	exo
33	0.00	1.00	1.00	0.00	endo
34	0.99	0.01	0.01	0.99	exo
35	0.01	0.99	0.99	0.01	endo
36	0.00	1.00	1.00	0.00	endo
37	0.94	0.06	0.06	0.94	exo
38	0.04	0.96	0.96	0.04	endo

\*=error

(点線以下はネットワークによる予測値)

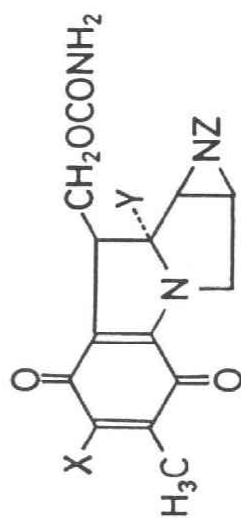
合物データに対する予測では、数値の混入がみられるが、endo および exo の教師ベクトルの各要素の数値を比較してコンホメーションを判断した（表右端）。その結果、1つの化合物（No.31）についてのみ誤った判断がなされた。したがって正答率は、 $12/13 = 0.92$ であり、線形学習機械法およびクラスター分析法の  $11/13 = 0.85$  より良い結果が得られた。<sup>24)</sup>

#### 4. 1. 2 マイトマイシン誘導体の等級活性分類

森口<sup>16)</sup>らは自ら開発した等級分類法である適応最小2乗（ALS）法を用いマイトマイシン誘導体の等級活性分類を行っている。この結果と比較するため、文献より同じデータを引用し、ニューラルネットワークに適用した。ネットワーク構成を  $N(6,12,5)$  とし、第2層のニューロンの  $\alpha$  値を 2.5、第3層のそれを 5.0 に設定した。 $\epsilon$  及び  $\theta$  は 0.1 及び 0.0 に設定した。入力データ種は  $F_x$ ,  $\delta_{m-x}$ ,  $V_{w-x}$ ,  $Y_{O-Me}$ ,  $Y_{OH}$ , 及び  $E_{s-z}$  であり、それらの値を 0.1 と 1.0 の間にスケールして用いた。ここで、最小値を 0.1 とした理由は、0 に設定すると、その値を有する第1, 2層間の結合荷重の学習が不可能になることを避けるためである。抗ガン活性の5段階（3+ ~ -）の等級に対し5個の要素を有する教師ベクトルを対応させた。例えば、+2 のランクに対する教師パターンを (0,1,0,0,0) とした。表4-3 に入力データおよび出力ベクトルを示す、表の右端のカラムに観測された生物活性等級と、ネットワークの等級分類を示す、すべての誘導体は正しく分類された。ALS法では正しく分類されない化合物が1個ある。ニューラルネットワーク法は、この場合より高い認識能力を示した。

ALS法に比べて、ニューラルネットワーク法のすぐれた点は、未知データについて予測可能な点である。次にこの能力を示すため、マイトマイシン誘導体1~16のうち、任意の5つの化合物（2, 4, 7, 10 及び 16）を除いて学習させ、次に除いた化合物を未知化合物としてネットワークに与え、活性分類を予測させた。

表4-3. マイトマイシンの構造, 生理活性等級と  
ニューラルネットワークの入力データ, 計算された等級



No.	X	Y	Z	F <sub>x</sub>	σ <sub>m-x</sub>	V <sub>w-x</sub>	Y <sub>OMe</sub>	Y <sub>OH</sub>	E <sub>s-z</sub>	rank	output pattern				obs.	calcd.
											2+	+	+/-	-		
1	NH <sub>2</sub>	OMe	H	0.02	-0.16	0.177	1	0	1.24	0.95	0.08	0.00	0.00	0.00	0.00	3+
2	NHET	OMe	H	-0.11	-0.24	0.493	1	0	1.24	0.95	0.05	0.00	0.00	0.00	0.00	3+
3	NH <sub>2</sub>	OMe	Me	0.02	-0.16	0.177	1	0	0	0.01	0.98	0.06	0.00	0.00	0.00	2+
4	NH <sub>2</sub>	OMe	Et	0.02	-0.16	0.177	1	0	-0.07	0.00	0.98	0.10	0.00	0.00	0.00	2+
5	NH <sub>2</sub>	OMe	Ac	0.02	-0.16	0.177	1	0	-0.47	0.03	0.94	0.56	0.00	0.00	0.00	2+
6	NH <sub>2</sub>	OH	Me	0.02	-0.16	0.177	0	1	0	0.03	1.00	0.00	0.02	0.00	0.00	2+
7	NMe <sub>2</sub>	OMe	H	0.10	-0.15	0.441	1	0	1.24	0.02	0.92	0.03	0.00	0.00	0.00	2+
8	NH <sub>2</sub>	OMe	COPh-o-Cl	0.02	-0.16	0.177	1	0	-1.19	0.00	0.04	0.84	0.00	0.07	+	+
9	NH <sub>2</sub>	OMe	COPh-p-Cl	0.02	-0.16	0.177	1	0	-1.19	0.00	0.04	0.84	0.00	0.07	+	+
10	NHPh	OMe	H	-0.02	-0.12	0.892	1	0	1.24	0.00	0.08	0.96	0.00	0.00	+	+
11	OMe	OMe	H	0.26	0.12	0.304	1	0	1.24	0.00	0.08	0.99	0.00	0.00	+	+
12	OMe	OMe	Me	0.26	0.12	0.304	1	0	0	0.00	0.07	1.00	0.00	0.00	+	+
13	OMe	OH	Me	0.26	0.12	0.304	0	1	0	0.00	0.05	0.01	0.96	0.03	+/-	+/-
14	NH <sub>2</sub>	H	Me	0.02	-0.16	0.177	0	0	0	0.00	0.00	0.00	0.00	0.94	-	-
15	NH <sub>2</sub>	OMe	SO <sub>2</sub> Me	0.02	-0.16	0.177	1	0	-1.54	0.00	0.00	0.45	0.03	0.89	-	-
16	OMe	H	Me	0.26	0.12	0.304	0	0	0	0.00	0.00	0.06	0.00	0.99	-	-

Me:CH<sub>3</sub>, Et:C<sub>2</sub>H<sub>5</sub>, Ac:CH<sub>3</sub>CO, Ph:C<sub>6</sub>H<sub>5</sub>

ネットワークの構造とパラメータは前と同じとした。結果を表4-4 に示す。正しく等級分類されている未知化合物に対する予測では化合物7 と 10 に対しそれぞれ+1, +2 等級の誤りがあるが、その誤りは軽微であるといえる。<sup>25)</sup>

#### 4. 1. 3 アリルアクリロイルピペラジンの等級活性分類

アリルアクリロイルピペラジン誘導体は抗高血圧の生物活性を有する。関谷<sup>23)</sup>らは、これらの化合物に対しALS法を適応し等級分類を試みたが、結果は良好ではなかった。この問題にニューラルネットワーク法の適用を試みた。比較のため、化合物番号及び入力データは文献と同じものを用いた。入力データ種は $\Delta RI$ ,  $\Sigma x$ ,  $\Sigma \sigma$ ,  $I(2-OR)$ ,  $I(RI)$ , 及びHBの6種である。ネットワーク構造を  $N(7,14,4)$  とした、入力層のニューロン数を7個としたのは、常に定数1の要素を加えるためである。

予想されたとおり学習の収束は困難であった。784回目の学習でのE値は1.0であった。中間層のニューロン数を2倍、すなわち  $N(7,28,4)$  とし、学習の回数を1410回としてもE値は1.0以下にならなかった。これは入力データの中に矛盾を含むためである。表4-5 に入力データ及び出力データを示す。収束が完全でなかったが、殆どの化合物は正しく等級分類されている。誤りは化合物番号34, 35, 及び38においてみられたが、それらの出力ベクトルの要素は興味深い内容を含んでいる。すなわち化合物34 及び 35 の出力ベクトルはそれぞれ(0.00,0.51, 0.54,0.00) 及び(0.00,0.60,0.49,0.00) であり、両者ともランク2 及び 3 に対する要素に大きな数値が入っている。これは化合物34 では、教師パターンはランク2 を提示しているが、ネットワークは、他のデータの化合物から得られた情報からランク3を主張し、一方化合物35 では、ランク3の提示に対し、ネットワークはランク2を主張していることを示唆している。このように出力ベクトルを解析すると、データの中にある矛盾あるいは測定誤りに関する情報を得ることが

表4-4. マイトマイシンのSAR予測：

学習データ，未学習データ（点線以下）の出力パターンと等級

compd. No	rank	output pattern					obs. class./pred.	
		3+	2+	+	+/-	-		
1	0.91	0.13	0.00	0.00	0.00	0.00	3+	3+
3	0.10	0.87	0.05	0.00	0.00	0.00	2+	2+
5	0.01	0.93	0.32	0.00	0.00	0.00	2+	2+
6	0.00	0.98	0.00	0.02	0.00	0.00	2+	2+
8	0.00	0.12	0.48	0.01	0.00	0.00	+	+
9	0.00	0.12	0.48	0.01	0.00	0.00	+	+
11	0.00	0.04	0.92	0.00	0.00	0.00	+	+
12	0.00	0.00	0.96	0.01	0.04	0.04	+	+
13	0.00	0.00	0.02	0.93	0.05	0.05	+/-	+/-
14	0.00	0.02	0.00	0.00	0.91	0.91	-	-
15	0.00	0.00	0.35	0.06	0.87	0.87	-	-
<hr/>								
2	0.91	0.14	0.00	0.00	0.00	0.00	3+	3+
4	0.07	0.89	0.07	0.00	0.00	0.00	2+	2+
7	0.81	0.25	0.00	0.00	0.00	0.00	2+	3+
10	0.84	0.21	0.00	0.00	0.00	0.00	+	3+
16	0.00	0.00	0.74	0.00	0.99	0.99	-	-



表4-5. アリルアクリロイルピペラジンのSAR:  
入力データと計算された等級

No	R1	R <sup>2</sup>	ΔRI	Σ π	Σ σ	I(2-OR)	I(R <sup>1</sup> )	HB	rank 1	output pattern			obs.	calcd.
										2	3	4		
19	H	4-i-PrPh	2.36	1.53	-0.15	0	0	0	0	0.989	0.090	0.001	0.000	1
20	H	2-MeOPh	0.49	-0.02	-0.27	1	0	1	0	0.989	0.159	0.000	0.000	1
21	H	2-EtOPh	1.18	0.38	-0.24	1	0	1	0	1.000	0.056	0.000	0.000	1
22	H	3,5-diMeOPh	0.56	-0.04	0.24	0	0	1	0	0.999	0.000	0.000	0.005	1
23	H	3-NO <sub>2</sub> Ph	0.03	-0.28	0.71	0	0	1	0	1.000	0.000	0.000	0.000	1
24	H	3,4-diClPh	2.19	1.42	0.60	0	0	0	0	1.000	0.001	0.000	0.000	1
25	H	3-CF <sub>3</sub> Ph	1.17	0.88	0.43	0	0	0	0	1.000	0.008	0.000	0.000	1
26	H	3-MeOPh	0.28	-0.02	0.12	0	0	1	0	0.809	0.000	0.001	0.097	1
27	H	4-ClPh	1.01	0.71	0.23	0	0	0	0	0.992	0.073	0.004	0.000	1
28	H	4-BrPh	1.26	0.86	0.23	0	0	0	0	0.998	0.045	0.001	0.000	1
29	H	5-Cl-2-Th	0.89	0.39	0.26	0	0	0	0	0.982	0.081	0.009	0.000	1
30	H	3-MePh	1.01	0.56	-0.07	0	0	0	0	0.104	0.482	0.199	0.000	2
31	H	4-MePh	0.95	0.56	-0.17	0	0	0	0	0.009	0.621	0.425	0.000	2
32	H	4-i-PrOPh	1.53	0.82	-0.45	0	0	1	0	0.019	0.887	0.000	0.796	2
33	H	2,3,4-triMeOPh	-0.09	-0.06	-0.42	1	0	1	0	0.002	0.858	0.004	0.029	2
34	H	5-Me-2-Th	0.69	0.24	-0.14	0	0	0	0	0.002	0.513	0.537	0.001	2
35	H	2-MePh	0.83	0.56	-0.17	0	0	0	0	0.005	0.597	0.485	0.000	3
36	Me	Ph	0.20	0.54	0	0	1	0	0	0.000	0.042	0.882	0.002	3
37	H	3-Fu	-1.29	-0.79	0.04	0	0	1	0	0.000	0.000	0.881	1.000	3
38	H	5-Me-2-Fu	0.04	-0.23	0.15	0	0	1	0	0.111	0.000	0.003	0.137	3
39	Me	2-Th	-0.13	0.22	0.03	0	1	0	0	0.000	0.003	0.914	0.051	3
40	H	3-Me-2-Th	0.48	0.24	-0.14	0	0	0	0	0.001	0.401	0.606	0.004	3
18	H	Ph	0	0	0	0	0	0	0	0.000	0.077	0.538	0.023	3
41	H	2-Th	-0.36	-0.32	0.03	0	0	0	0	0.000	0.005	0.565	0.396	3
42	H	4-MeOPh	0.19	-0.02	-0.27	0	0	1	0	0.000	0.030	0.070	0.993	4
43	H	4-EtOPh	0.92	0.38	-0.24	0	0	1	0	0.091	0.121	0.001	0.578	4
44	H	2-Fu	-0.84	-0.79	0.32	0	0	1	0	0.000	0.000	0.061	0.886	4
45	Me	2-Fu	-0.71	-0.25	0.32	0	1	1	0	0.000	0.000	0.071	0.989	4
46	H	3-Th	-0.61	-0.32	0.04	0	0	0	0	0.000	0.001	0.577	0.834	4

Ph: C<sub>6</sub>H<sub>5</sub>, Fu: Furyl (C<sub>4</sub>H<sub>3</sub>O), Th: Thienyl (C<sub>4</sub>H<sub>3</sub>S)

できる。

アリルアクリロイルピペラジンの生物活性データには矛盾を含むことがわかったが、同じデータを用いてニューラルネットワークの予測能力を調べた。8個の化合物に関するデータを除いて学習させ、除いた8個の化合物を未知データとしてネットワークに与え予測させた。結果を表4-6に示す。分類の正答率は $19/21 = 0.90$ であり、未学習化合物に対する予測の正答率は $6/8 = 0.75$ であった。<sup>25)</sup>

以上みてきたように、ニューラルネットワークの基本的分類動作を薬物の等級活性分類に適用したところ、現在もっとも信頼されているALS法より優れている可能性が示唆された。これは第2章で示したニューラルネットワーク法がALS法の近似を拡張していることが原因かもしれない。またALS法にない利点は、未知化合物の生物活性強度を予測できることである。

## 4. 2 定量的構造活性相関研究への応用

第2章において述べたように、MR型ニューラルネットワークは非線形重回帰分析法 (nonlinear multiregression analysis) に近縁の方法である。そして特別な場合として重回帰分析法の近似に変形できる。QSAR研究では重回帰分析法が用いられている。本節ではMR型ニューラルネットワーク法をQSARに適用し従来法の結果との比較を行う。

### 4. 2. 1 カルボキノン誘導体のQSAR

カルボキノン誘導体は制ガン抗生物質マイトマイシンCに基づくデザインから、中尾ら<sup>26, 27, 35, 36, 37)</sup>によって合成された抗ガン活性を有する一連の化合物で、実際に医療に用いられている。Hansch法 (線形重回帰分析法)<sup>28, 29, 30)</sup>を用い

表4-6. アリルアクリロイルピペラジンのSAR予測:

学習データ, 未学習データ (点線以下) の出力パターンと等級

No	output pattern				class./pred.
	rank 1	2	3	4	
19	1.000	0.185	0.001	0.000	1
21	0.999	0.007	0.000	0.000	1
23	0.999	0.000	0.011	0.019	1
25	1.000	0.015	0.002	0.000	1
26	0.889	0.000	0.410	0.045	1
27	0.999	0.068	0.010	0.000	1
29	0.987	0.067	0.022	0.000	1
30	0.090	0.403	0.166	0.000	2
31	0.004	0.551	0.306	0.000	2
32	0.142	0.779	0.000	0.041	2
34	0.000	0.476	0.609	0.000	3
35	0.002	0.533	0.377	0.000	2
36	0.000	0.023	0.964	0.001	3
37	0.000	0.000	0.843	0.213	3
38	0.324	0.000	0.589	0.134	3
40	0.000	0.403	0.702	0.001	3
41	0.000	0.006	0.601	0.326	3
42	0.000	0.005	0.082	0.961	4
43	0.038	0.105	0.003	0.509	4
44	0.000	0.000	0.181	0.796	4
46	0.000	0.002	0.512	0.576	4
-----					
20	0.569	0.001	0.029	0.101	1
22	0.999	0.000	0.222	0.003	1
24	1.000	0.005	0.000	0.000	1
28	1.000	0.059	0.005	0.000	1
33	0.000	0.000	0.159	0.715	4
39	0.000	0.002	0.981	0.030	3
18	0.000	0.074	0.694	0.022	3
45	0.000	0.000	0.974	0.017	3

た詳しいQ S A R研究は吉本ら<sup>31)</sup>によって行われている。はじめにこの問題にMR型ニューラルネットワーク法を適用した。

入力データは分子屈折定数(MR)，疎水性定数( $\pi$ )，置換基定数(F及びR)，及び立体効果と疎水性定数を推定するための $MR_{1,2}$ 及び $\pi_{1,2}$ の6種である。教師データとなる生物活性値はL-1210白血球細胞を接種したマウスの40%延命効果をあらわす薬量(minimum effective dose: MED)，最大延命効果を表す薬量(optimal dose: OD)のmol/kgをCとし， $\log 1/C$ を用いた。投薬は少量ずつ連続的に投与する方法(chronic injection)と1回だけ投与する方法(single injection)が用いられている。

ネットワーク構造による結果の違いを調べるため， $N(7,12,1)$ 及び $N(13,26,1)$ を同じ問題へ適用した。すなわち，後者の構造には構造パラメータとそれらの2乗値， $MR_{1,2}^2$ ， $\pi_{1,2}^2$ ， $\pi^2$ ， $MR_1^2$ ， $F^2$ 及び $R^2$ ，を取り入れ結果に与える影響を調べた。ネットワークパラメータを表4-7に，また，構造データを表4-8に示す。表4-7Bにおいて，学習速度を速めるため $\alpha$ 値を5.0とした。結果を表4-9に示す。表の中で，clc 1, 2, 3, 及び 4 (それぞれ，カラム 3, 7, 11, 及び 15 に対応する)は文献<sup>31)</sup>より引用した従来方法で計算された結果であり，set A 及び set B はそれぞれ，ネットワーク A 及び B による結果である。+，±，及び-で示した行にはニューラルネットワークの結果が従来法の結果に比べてそれぞれ，良かった(測定値により近い)，同程度の，及び悪かったケースの数である。ニューラルネットワーク法は重回帰分析法より，良い結果を与える確率が高い。2乗項の取り込みにより結果はより改善され，入力データ種の数が増えても不都合が生じない。

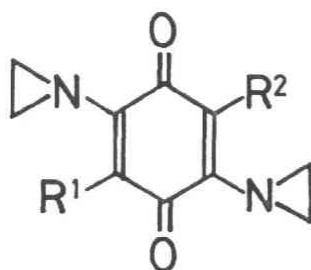
表4-10 にニューラルネットワーク法と重回帰分析法との平均偏差(MD)，分散(variance)，及び標準偏差(SD)を示す。この表からわかるようにニューラルネットワーク法は全ての場合において重回帰分析法を凌いでいる。ニューラル

表4-7. ニューラルネットワークの構造とパラメータ

	ly <sup>a</sup>	nnr <sup>b</sup>	$\alpha$	$\varepsilon$	$\theta$
A	1	7			
	2	12	2	0.10	0
	3	1	-	0.05	-
-----					
B	1	13			
	2	26	5	0.05	0
	3	1	-	0.05	-
-----					
C	1	14			
	2	28	5	0.05	0
	3	1	-	0.05	-
-----					

<sup>a</sup>Layer, <sup>b</sup>Number of neurons.

表4-8. カルボキノンの入力データ



No.			MR <sub>1,2</sub>	π <sub>1,2</sub>	π <sub>2</sub>	MR <sub>1</sub>	F	R
1	CH <sub>3</sub>	COCH <sub>3</sub>	1.69	-0.05	-0.55	0.57	0.28	0.07
2	C <sub>6</sub> H <sub>5</sub>	C <sub>6</sub> H <sub>5</sub>	5.08	3.92	1.96	2.54	0.16	-0.16
3	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>3</sub> C <sub>6</sub> H <sub>5</sub>	4.50	3.66	3.16	0.57	-0.08	-0.26
4	C <sub>5</sub> H <sub>11</sub>	C <sub>5</sub> H <sub>11</sub>	4.86	5.00	2.50	2.43	-0.08	-0.26
5	CH(CH <sub>3</sub> ) <sub>2</sub>	CH(CH <sub>3</sub> ) <sub>2</sub>	3.00	2.60	1.30	1.50	-0.08	-0.26
6	CH <sub>3</sub>	CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	3.57	2.51	2.01	0.57	-0.12	-0.14
7	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	3.00	3.00	1.50	1.50	-0.08	-0.26
8	CH <sub>3</sub>	CH <sub>2</sub> OC <sub>6</sub> H <sub>5</sub>	3.79	2.16	1.66	0.57	-0.04	-0.13
9	R <sup>1</sup> =R <sup>2</sup> =CH <sub>2</sub> CH <sub>2</sub> OCON(CH <sub>3</sub> ) <sub>2</sub>		6.14	0.72	0.36	3.07	-0.08	-0.26
10	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	2.06	2.00	1.00	1.03	-0.08	-0.26
11	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	2.28	1.03	0.53	0.57	-0.08	-0.26
12	OCH <sub>3</sub>	OCH <sub>3</sub>	1.58	-0.04	-0.02	0.79	0.52	-1.02
13	CH <sub>3</sub>	CH(CH <sub>3</sub> ) <sub>2</sub>	2.07	1.80	1.30	0.57	-0.08	-0.26
14	C <sub>3</sub> H <sub>7</sub>	CH(OCH <sub>3</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	4.24	0.98	-0.52	1.50	-0.04	-0.13
15	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCON(CH <sub>3</sub> ) <sub>2</sub>	3.64	0.86	0.36	0.57	-0.08	-0.26
16	CH <sub>3</sub>	CH <sub>3</sub>	1.14	1.00	0.50	0.57	-0.08	-0.26
17	H	CH(CH <sub>3</sub> ) <sub>2</sub>	1.60	1.30	1.30	0.10	-0.04	-0.13
18	CH <sub>3</sub>	CH(OCH <sub>3</sub> )C <sub>2</sub> H <sub>5</sub>	2.75	1.53	1.03	0.57	-0.04	-0.13
19	C <sub>3</sub> H <sub>7</sub>	CH <sub>2</sub> CH <sub>2</sub> OCONH <sub>2</sub>	3.56	1.45	-0.05	1.50	-0.08	-0.26
20	R <sup>1</sup> =R <sup>2</sup> =CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>		3.42	1.03	0.53	1.71	-0.08	-0.26
21	C <sub>2</sub> H <sub>5</sub>	CH(OC <sub>2</sub> H <sub>5</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	4.23	0.98	-0.02	1.03	-0.04	-0.13
22	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCOCH <sub>3</sub>	2.78	1.23	0.73	0.57	-0.08	-0.26
23	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>3</sub> -dimer	1.96	2.00	1.50	0.57	-0.08	-0.26
24	CH <sub>3</sub>	C <sub>2</sub> H <sub>5</sub>	1.60	1.50	1.00	0.57	-0.08	-0.26

25	CH <sub>3</sub>	CH(OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub> )- CH <sub>2</sub> OCONH <sub>2</sub>	4.45	0.01	-0.49	0.57	-0.04	-0.13
26	CH <sub>3</sub>	CH <sub>2</sub> CH(CH <sub>3</sub> )OCONH <sub>2</sub>	3.09	0.75	0.25	0.57	-0.08	-0.26
27	C <sub>2</sub> H <sub>5</sub>	CH(OCH <sub>3</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	3.77	0.48	-0.52	1.03	-0.04	-0.13
28	CH <sub>3</sub>	CH(C <sub>2</sub> H <sub>5</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	3.55	1.25	0.75	0.57	-0.08	-0.26
29	CH <sub>3</sub>	CH(OC <sub>2</sub> H <sub>5</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	3.77	0.48	-0.02	0.57	-0.04	-0.13
30	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>3</sub> OCONH <sub>2</sub>	3.09	0.95	0.45	0.57	-0.08	-0.26
31	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>2</sub> OCONH <sub>2</sub>	2.63	0.45	-0.05	0.57	-0.08	-0.26
32	C <sub>2</sub> H <sub>5</sub>	(CH <sub>2</sub> ) <sub>2</sub> OCONH <sub>2</sub>	3.09	0.95	-0.05	1.03	-0.08	-0.26
33	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OH	1.78	0.34	-0.16	0.57	-0.08	-0.26
34	CH <sub>3</sub>	CH(CH <sub>3</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	3.09	0.75	0.25	0.57	-0.08	-0.26
35	CH <sub>3</sub>	CH(OCH <sub>3</sub> )CH <sub>2</sub> OCONH <sub>2</sub>	3.31	-0.02	-0.52	0.57	-0.04	-0.13
36	H	N(CH <sub>2</sub> ) <sub>2</sub>	1.66	0.18	0.18	0.10	0.10	-0.92
37	R <sup>1</sup> =R <sup>2</sup> =CH <sub>2</sub> CH <sub>2</sub> OH		2.42	-0.32	-0.16	1.21	-0.08	-0.26
38	CH <sub>3</sub>	N(CH <sub>2</sub> ) <sub>2</sub>	2.13	0.68	0.18	0.57	0.06	-1.05
39	CH <sub>3</sub>	CH(OCH <sub>3</sub> )CH <sub>2</sub> OH	2.47	-0.13	-0.63	0.57	-0.04	-0.13

\*The data were taken from the literature (Ref.31).

表4-9. ニューラルネットワーク，重回帰分析の結果比較

No	chron. inj., log(1/C)								sing. inj., log(1/C)							
	MED				OD				MED				OD			
	obsd. <sup>a</sup>	clc.1 <sup>b</sup>	set A <sup>c</sup>	set B <sup>d</sup>	obsd. <sup>a</sup>	clc.2 <sup>b</sup>	set A <sup>c</sup>	set B <sup>d</sup>	obsd. <sup>a</sup>	clc.3 <sup>b</sup>	set A <sup>c</sup>	set B <sup>d</sup>	obsd. <sup>a</sup>	clc.4 <sup>b</sup>	set A <sup>c</sup>	set B <sup>d</sup>
1									3.94	4.12	4.260	4.387	3.48	3.72	3.757	3.582
2	4.33	4.05	4.260	4.362	4.14	3.81	4.062	4.155					3.53	2.97	3.301	3.426
3	4.47	4.61	4.629	4.549	4.21	4.57	4.561	4.452	3.93	4.23	4.268	4.178	3.60	3.99	3.923	3.717
4	4.63	4.31	4.327	4.297	4.52	4.27	4.211	4.167	4.07	3.58	3.825	3.840	3.62	3.50	3.615	3.491
5	4.77	5.26	5.097	5.142	4.59	4.96	4.899	4.920	4.36	4.74	4.658	4.624	4.14	4.38	4.292	4.195
6	4.85	5.18	5.109	5.087	4.69	4.92	4.944	4.874	4.74	4.77	4.746	4.672	4.26	4.37	4.320	4.169
7	4.92	5.15	4.976	5.006	4.44	4.89	4.802	4.804	4.32	4.55	4.530	4.488	4.14	4.23	4.171	4.047
8	5.15	5.21	5.132	5.206	4.71	4.87	4.893	4.916	4.68	4.61	4.611	4.692	3.89	4.23	4.167	4.123
9	5.16	5.21	5.193	5.048	4.85	4.83	4.862	4.815					4.62	5.07	4.647	4.537
10	5.46	5.57	5.463	5.525	5.09	5.20	5.227	5.253	4.94	5.01	4.997	4.992	4.79	4.58	4.575	4.489
11	5.57	5.98	5.977	6.021	5.42	5.50	5.621	5.657	5.19	5.51	5.524	5.579	5.12	4.95	5.019	4.998
12	5.59	5.74	5.707	5.729	5.17	5.27	5.294	5.264	4.81	4.79	4.559	4.629	4.32	4.45	4.144	4.196
13	5.60	5.58	5.536	5.579	5.21	5.23	5.325	5.339	4.96	5.13	5.100	5.108	4.69	4.67	4.657	4.616
14	5.63	6.03	5.797	5.808	5.07	5.37	5.229	5.239	5.01	5.18	5.191	5.117	4.64	4.66	4.645	4.446
15									5.09	5.59	5.626	5.684	4.84	5.01	5.068	5.023
16	5.66	5.99	6.067	6.088	5.36	5.51	5.742	5.730	5.36	5.52	5.539	5.505	4.79	4.96	5.067	4.971
17	5.68	5.56	5.657	5.728	5.37	5.13	5.393	5.388	5.16	5.02	5.024	5.032	4.59	4.54	4.548	4.564
18	5.68	5.54	5.498	5.617	5.33	5.09	5.186	5.236	5.26	4.91	4.907	5.014	4.84	4.46	4.430	4.429
19	5.68	5.96	5.720	5.736	5.23	5.43	5.291	5.329	4.90	5.30	5.286	5.214	4.42	4.80	4.792	4.650
20	5.69	5.59	5.564	5.611	5.31	5.17	5.247	5.298	5.18	5.51	5.157	5.089	4.71	4.95	4.799	4.786
21	5.76	5.93	5.763	5.829	5.24	5.33	5.253	5.294	5.40	5.18	5.195	5.241	4.64	4.66	4.641	4.515
22	5.78	5.87	5.824	5.874	5.78	5.43	5.492	5.539								
23	5.82	5.47	5.434	5.468	5.39	5.16	5.257	5.256								
24	5.86	5.73	5.732	5.776	5.37	5.33	5.485	5.495	5.16	5.28	5.259	5.255	4.52	4.78	4.807	4.755
25	6.03	6.33	6.326	6.272	5.39	5.62	5.649	5.613	5.45	5.65	5.751	5.739	4.96	5.01	5.096	4.891
26	6.14	6.12	6.104	6.125	5.79	5.60	5.667	5.706	5.86	5.64	5.690	5.759	5.18	5.05	5.140	5.106
27	6.16	6.19	6.100	6.115	5.22	5.50	5.488	5.494	5.62	5.42	5.471	5.487	4.92	4.84	4.885	4.712



28	6.18	5.86	5.767	5.805	5.66	5.42	5.417	5.467	6.03	5.40	5.402	5.462	5.20	4.87	4.877	4.850
29	6.18	6.09	6.061	6.113	5.22	5.46	5.512	5.541	5.53	5.42	5.472	5.564	4.62	4.84	4.879	4.805
30	6.18	6.02	5.977	6.013	5.93	5.53	5.581	5.628	5.55	5.55	5.573	5.644	5.48	4.98	5.037	5.017
31	6.21	6.28	6.333	6.323	5.75	5.50	5.844	5.857	5.83	5.79	5.871	5.921	5.46	5.16	5.313	5.249
32	6.25	6.12	6.016	6.038	5.48	5.57	5.557	5.596	5.98	5.55	5.571	5.591	4.88	4.98	5.043	4.940
33	6.39	6.34	6.466	6.433	5.79	5.74	5.980	5.956	5.89	5.84	5.938	5.935	5.25	5.20	5.399	5.287
34	6.41	6.12	6.104	6.125	5.71	5.60	5.667	5.706	5.93	5.64	5.690	5.759	5.31	5.05	5.140	5.106
35	6.41	6.35	6.424	6.404	5.66	5.64	5.770	5.744	5.81	5.67	5.767	5.813	5.03	5.03	5.145	5.001
36	6.45	6.54	6.623	6.581	6.19	6.16	6.300	6.319	6.02	6.19	6.291	6.287	5.74	5.60	5.750	5.663
37	6.54	6.12	6.376	6.333	6.05	5.56	5.907	5.875	5.93	6.16	5.894	5.852	5.60	5.45	5.494	5.429
38	6.77	6.56	6.468	6.426	6.21	6.25	6.210	6.194	6.54	6.30	6.376	6.310	5.69	5.72	5.878	5.638
39	6.90	6.40	6.559	6.522	5.75	5.67	5.903	5.848	6.05	5.72	5.833	5.851	5.27	5.07	5.228	5.060
-----																
+			17	19			21	22			17	24			22	24
±			3	0			1	0			6	1			1	1
-			17	18			15	15			12	10			14	12
-----																

\*Data by Yoshimoto et al.<sup>37)</sup>

<sup>b</sup>The results calculated by eqs 8, 10, 12, and 14 in Ref. 37.

<sup>c</sup>Results by the neural network with the parameter set A (Table 4-7 A).

<sup>d</sup>Results by the neural network with the parameter set B (Table 4-7 B).

+, ±, and - show the numbers of cases

in which the neural network is superior, equivalent, and inferior to the multi-regression analysis, respectively.

表4-10. 平均値, 分散, 標準偏差の比較

	連続投与						一回投与					
	MED			OD			MED			OD		
	set A <sup>a</sup>	set B <sup>b</sup>	MR <sup>c</sup>	set A <sup>a</sup>	set B <sup>b</sup>	MR <sup>c</sup>	set A <sup>a</sup>	set B <sup>b</sup>	MR <sup>c</sup>	set A <sup>a</sup>	set B <sup>b</sup>	MR <sup>c</sup>
MD	0.17	0.16	0.20	0.15	0.14	0.19	0.20	0.19	0.23	0.16	0.14	0.20
variance	0.044	0.044	0.059	0.036	0.032	0.054	0.064	0.058	0.073	0.040	0.030	0.059
SD	0.21	0.21	0.24	0.19	0.18	0.23	0.25	0.24	0.27	0.20	0.17	0.24

<sup>a</sup>Results by the parameter set A (Table 4-7 A).

<sup>b</sup>Results by the parameter set B (Table 4-7 B).

<sup>c</sup>Results by the multiregression analysis.

ネットワーク法に対する重回帰分析法の分散比 ( $F$ ) は, set A で  $1.14 \leq F \leq 1.50$ , set B では  $1.26 \leq F \leq 1.97$  であった。回帰係数の数は 6 であるので, これらの比は有効である。

つぎに, leave-n-out 法 ( $n = 1, 2, 5, 10$ ) を用いて set A での学習の汎化状況を調べた。すなわち, 37 個のデータから無作為に  $n$  個のデータを除いた  $37-n$  個のデータで学習し, 除いたデータを予測され, 平均偏差を求める。この操作を 37 回繰り返し, 偏差 ( $\sigma^2$ ) を平均する。結果を表4-11 に比較のため学習に用いたデータの偏差 ( $\sigma_L^2$ ) とともに示す。

未学習データに対する偏差は学習データに対するそれより大きい, 観測値をよく再現していると言える。また, 分散値が小さいことからここに用いられているデータは線形性が強いと言える。

#### 4. 2. 2 ベンゾジアゼピン誘導体の Q S A R

Randall ら<sup>32, 33)</sup> はベンゾジアゼピン誘導体の一つのクロロジアゼボキサイドをマイナートランキライザーとしてはじめて用いた。その後 1,4-ベンゾジアゼピン類の生物活性が徹底的に調べられ, 今日のマイナートランキライザーの分野において主要な分野を占めるに至った。窪田ら<sup>34)</sup>はこの系列の化合物の Q S A R 解析を行っている。本節では MR 型ニューラルネットワークを適用し, 窪田ら<sup>34)</sup>の結果と比較する。

構造パラメータ及び生物活性値は文献から引用しそのまま用いた。入力データは MR-3,  $\pi$ -3, MR-7,  $\sigma_m$ -3, F-4, R-4, I-1, 及び MR-3,  $\pi$ -3, MR-7,  $\sigma_m$ -3, F4, 及び R-4 の 2 乗及び 1 の定数項を用いた。構造パラメータについての番号 (X-3 etc) は表4-12 に示した構造の位置番号を示す。入力データは 0.1~1.0 の範囲に, 教師データは 0.0~1.0 にスケールした。各層のニューロン数及びバックプロパゲーションパラメータを表4-7C に示す。

表4-11. カルボキノンの Leave-n-out 法の分散値の平均\*

n	$\sigma^2$	$\sigma_L^2$
1 <sup>b</sup>	0.069	0.043
2	0.016	0.043
5	0.044	0.041
10	0.059	0.040

\*Allied to MED of chronic injection.

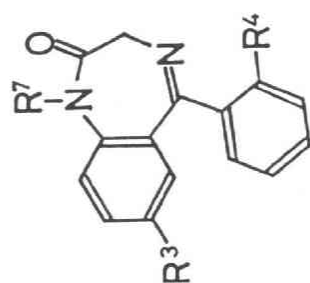
<sup>b</sup>Simple mean value of differences (observed value - calculated value).

入力データ，結果，及び文献値を表4-12 に示す。ニューラルネットワーク法と重回帰分析法との結果を比較すると，96 のケースについては良く，62 のケースは悪く，また 5 のケースは同等となった。表4-12 の信頼性を調べるため，平均偏差，分散，及び標準偏差を求め，重回帰分析法との結果と比較した。結果を表4-13 に示す。サンプル数は 54 である。分散比（F 値：ニューラルネットワーク／重回帰分析）は  $1.89 \leq F \leq 3.75$  であつた。したがって，ニューラルネットワーク法は重回帰分析法を凌いでいる。

アンチペンチレンテトラゾール効果に対して leave-n-out 法を試みた。57 個のデータから無作為に  $n$  個のデータを除いた， $57-n$  個のデータを用いて学習し，除いた  $n$  個のデータを予測し分散を求めた。この操作を 57 回繰り返し平均し  $\sigma^2$  を求める。結果を学習に用いたデータの分散 ( $\sigma_L^2$ ) とともに表4-14 に示す。

ニューラルネットワークは非線形関係にあるデータに対しては非線形 fitting を行う。非線形の程度は与えられたデータによつて定まる。したがって，与えられたデータの非線形性の度合いが強く，それを表現するデータ数が十分多くないと予測のはずれる可能性が大きくなる。前例のカルボキノンの場合とは異なり，ベンゾジアゼピン類のデータの非線形性は強く  $\sigma^2$  値は一桁大きくなる。その結果重回帰分析の leave-n-out 法の結果を越えていない。

表4-12. ベンゾジアゼピン誘導体の入力データと出力値



No.	substituent	MR-3	$\pi$ -3	MR-7	$\sigma$ m-3	F-4	R-4	I-1	anti-pent. effect <sup>b</sup>		anti-fighting behavior		clined screen test	
									obsd. <sup>a</sup>	clc.1 <sup>c</sup>	obsd. <sup>a</sup>	clc.2 <sup>c</sup>	obsd. <sup>a</sup>	clc.3 <sup>c</sup>
1	3-Cl-7-iso-C <sub>6</sub> H <sub>11</sub>	0.60	0.71	2.42	0.37	0.0	0.0	0.0	4.99	4.76	4.876	3.78	2.83	3.15
2	3,4-F <sub>2</sub> -7-CH <sub>3</sub>	0.09	0.14	0.57	0.34	0.43	-0.34	0.0	3.33	6.40	3.447		2.85	2.856
3	3-SC <sub>2</sub> H <sub>5</sub>	1.84	1.07	0.10	0.15	0.0	0.0	1.0	3.57	4.00	3.874			2.658
4	3-Cl-7-CH <sub>2</sub> CONHCH <sub>3</sub>	0.60	0.71	1.92	0.37	0.0	0.0	0.0	5.83	4.95	5.012			2.932
5	3-SC <sub>4</sub> H <sub>9</sub>	2.77	2.07	0.10	0.15	0.0	0.0	1.0	3.79	3.50	3.659			
6	3-NO <sub>2</sub> -7-iso-C <sub>6</sub> H <sub>11</sub>	0.74	-0.28	2.42	0.71	0.0	0.0	0.0	4.80	5.02	5.027			2.691
7	3-N(CH <sub>3</sub> ) <sub>2</sub>	1.56	0.18	0.10	-0.15	0.0	0.0	1.0	3.84	3.85	3.994			
8	3-Cl-4-OCH <sub>3</sub>	0.60	0.71	0.10	0.37	0.26	-0.51	1.0	4.60	5.30	4.638			
9	3-Cl-7-(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.60	0.71	2.95	0.37	0.0	0.0	0.0	3.38	4.56	4.712			
10	3-Cl-7-(CH <sub>2</sub> ) <sub>3</sub> OH	0.60	0.71	1.65	0.37	0.0	0.0	0.0	4.34	5.05	5.081			
11	3-NO <sub>2</sub> -7-CH <sub>2</sub> CONHCH <sub>3</sub>	0.74	-0.28	1.92	0.71	0.0	0.0	0.0	5.29	5.21	5.289			
12	3-Cl-7-(CH <sub>2</sub> ) <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	0.60	0.71	3.41	0.37	0.0	0.0	0.0	5.06	4.39	4.545			
13	3-Cl-7-CH <sub>2</sub> CON(CH <sub>3</sub> ) <sub>2</sub>	0.60	0.71	2.39	0.37	0.0	0.0	0.0	5.35	4.77	4.884			
14	4-F-7-CH <sub>3</sub>	0.10	0.0	0.57	0.0	0.43	-0.34	0.0	4.53	6.06	4.359			
15	3-Cl-7-CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	0.60	0.71	3.00	0.37	0.0	0.0	0.0	4.64	4.54	4.686			
16	3-Cl-7-(CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.60	0.71	2.48	0.37	0.0	0.0	0.0	4.73	4.74	4.858			
17	3-Cl-4-F-7-(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.60	0.71	2.95	0.37	0.43	-0.34	0.0	4.76	4.27	5.193			
18	H	0.10	0.00	0.10	0.00	0.00	0.00	1.0						
19	3-CF <sub>3</sub> -7-CH <sub>2</sub> CONHCH <sub>3</sub>	0.50	0.88	1.92	0.43	0.00	0.00	0.0	5.06	5.06	5.088			
20	3-SC <sub>2</sub> H <sub>5</sub>	1.38	0.61	0.10	0.15	0.00	0.00	1.0	4.15	5.24	3.953			
21	3-Cl-7-(CH <sub>2</sub> ) <sub>3</sub> OH	0.60	0.71	1.65	0.37	0.00	0.00	0.0						
22	3-Cl-7-CH <sub>2</sub> CONH <sub>2</sub>	0.60	0.71	1.44	0.37	0.00	0.00	0.0	5.07	5.13	5.132			
23	3-F	0.09	0.14	0.10	0.34	0.00	0.00	1.0	4.08	4.61	4.338			
24	3-SOCH <sub>3</sub>	1.37	-1.58	0.10	0.52	0.00	0.00	1.0						

Table 4-12 to be continued.

25	3-Cl-4-CH <sub>3</sub>	0.60	0.71	0.10	0.37	-0.04	-0.13	1.0	4.57	4.81	5.115	3.45	4.28	3.727	3.28	3.36	3.583
26	3-(CH <sub>3</sub> ) <sub>2</sub> -7-CH <sub>3</sub>	1.56	0.18	0.57	-0.15	0.00	0.00	0.0	4.69	4.43	4.561	3.86	3.44	3.849	3.29	3.13	2.952
27	3-Cl-4-F-7-(CH <sub>2</sub> ) <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	0.60	0.71	3.41	0.37	0.43	-0.34	0.0	5.38	5.09	5.187	4.29	4.02	4.215	3.29	3.00	3.322
28	3-NO <sub>2</sub> -7-(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>3</sub> ) <sub>2</sub>	1.56	0.18	0.57	0.37	0.41	-0.15	0.0	5.82	5.10	6.100	3.61	3.80	3.608	3.01	3.02	2.792
29	3-NO <sub>2</sub> -7-(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.74	-0.28	2.95	-0.15	0.00	0.00	0.0	4.28	4.82	4.573	3.96	4.10	3.885	3.35	3.20	3.218
30	3-NO <sub>2</sub> -7-(CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.74	-0.28	2.48	0.71	0.00	0.00	0.0	4.70	5.00	5.002	4.25	4.26	3.938	3.37	3.40	2.680
31	3-Cl-4-Cl	0.60	0.71	0.10	0.37	0.41	-0.15	1.0	5.85	5.54	6.216	5.18	4.52	4.960	3.48	3.61	3.654
32	3-Cl-7-CH <sub>2</sub> -C≡C-C <sub>3</sub> H <sub>5</sub>	0.60	0.71	1.82	0.37	0.00	0.00	0.0	4.90	4.99	5.038	4.51	3.99	3.953	3.51	3.41	2.967
33	3-CN	0.63	-0.57	0.10	0.56	0.00	0.00	1.0	5.30	5.04	5.458	3.81	4.37	4.004	3.54	3.73	3.358
34	3-NO <sub>2</sub> -4-CF <sub>3</sub>	0.74	-0.28	0.10	0.71	0.38	0.19	1.0	5.70	5.76	5.835	4.54	4.64	4.371	3.54	3.93	3.229
35	3-Cl	0.60	0.71	0.10	0.56	0.00	0.00	1.0	4.65	4.87	5.118	4.75	4.94	4.663	3.56	3.73	3.924
36	3-CN-4-F	0.63	-0.57	0.10	0.56	0.43	-0.34	1.0	5.63	5.75	5.819	4.75	4.94	4.663	3.57	4.00	3.136
37	3-Cl-7-C <sub>2</sub> H <sub>5</sub>	0.60	0.71	1.03	0.37	0.00	0.00	0.0	4.90	5.28	5.226	4.17	4.25	4.350	3.60	3.74	3.549
38	3-SCH <sub>3</sub> -7-CH <sub>3</sub>	1.38	0.61	0.57	0.15	0.00	0.00	0.0	3.60	3.64	4.267	3.87	4.34	3.915	3.77	4.25	3.190
39	3-Cl-7-CH <sub>2</sub> COCH <sub>3</sub>	0.60	0.71	1.51	0.37	0.00	0.00	0.0	5.28	5.10	5.115	4.21	4.09	4.089	3.82	3.54	3.133
40	3-Cl-4-Br	0.60	0.71	0.10	0.37	0.44	-0.17	1.0	5.77	5.59	6.298	4.85	4.56	5.003	3.85	3.42	3.648
41	3-Cl-4-F	0.60	0.71	0.10	0.37	0.43	-0.34	1.0	6.46	5.58	6.112	4.76	4.73	4.846	3.86	3.99	3.483
42	3-NO <sub>2</sub> -4-CF <sub>3</sub> -7-CH <sub>3</sub>	0.74	-0.28	0.57	0.71	0.38	0.19	0.0	5.71	5.58	5.542	4.86	4.89	4.907	3.86	4.15	3.021
43	3-CF <sub>3</sub> -7-(CH <sub>2</sub> ) <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	0.50	0.88	2.48	0.43	0.00	0.00	0.0	4.76	4.85	4.939	4.27	3.89	3.897	3.88	3.24	2.902
44	3-Cl-7-CH <sub>2</sub> CH=CH <sub>2</sub>	0.60	0.71	1.45	0.37	0.00	0.00	0.0	5.35	5.13	5.130	4.49	4.11	4.119	3.89	3.56	3.175
45	3,4-Cl <sub>2</sub> -7-CH <sub>3</sub>	0.60	0.71	0.57	0.37	0.41	-0.15	0.0	6.03	6.13	5.929	4.50	4.76	4.400	3.90	3.82	3.297
46	3-Cl-7-CH <sub>3</sub>	0.60	0.71	0.57	0.37	0.00	0.00	0.0	5.31	5.46	5.313	4.45	4.41	4.606	3.98	3.94	4.043
47	3-NO <sub>2</sub> -4-Cl-7-CH <sub>3</sub>	0.74	-0.28	0.57	0.71	0.41	-0.15	0.0	6.92	6.39	6.732	5.12	5.26	5.210	4.04	4.09	3.618
48	3-CF <sub>3</sub> -4-CF <sub>3</sub>	0.50	0.88	0.10	0.43	0.38	0.19	1.0	4.97	5.61	5.239	4.27	4.27	4.400	4.09	3.78	3.992
49	3-Br	0.89	0.86	0.10	0.39	0.00	0.00	1.0	5.20	4.74	4.825	4.20	4.21	4.408	4.10	3.78	4.063
50	3-CN-7-CH <sub>3</sub>	0.63	-0.57	0.57	0.56	0.00	0.00	0.0	5.44	5.63	5.524	4.74	4.61	4.811	4.14	3.95	3.380
51	3-NO <sub>2</sub>	0.74	-0.28	0.10	0.71	0.00	0.00	1.0	5.60	5.13	5.652	4.75	4.66	4.568	4.27	4.00	4.152
52	3-NO <sub>2</sub> -7-CH <sub>3</sub>	0.74	-0.28	0.57	0.71	0.00	0.00	0.0	5.62	5.72	5.986	5.07	4.90	5.130	4.47	4.21	4.286
53	3-Cl	0.50	0.89	0.10	0.43	0.00	0.00	1.0	5.48	4.99	5.375	4.48	4.28	4.378	4.48	3.84	4.185
54	3-Cl-4-F-7-CH <sub>3</sub>	0.60	0.71	0.57	0.37	0.00	-0.34	0.0	5.88	5.46	5.806	4.78	4.77	4.823	4.48	3.94	4.071
55	3-CF <sub>3</sub> -7-CH <sub>2</sub> CH=CH <sub>2</sub>	0.50	0.89	1.45	0.43	0.00	0.00	0.0	5.03	5.24	5.211	4.54	4.23	4.382	4.54	3.68	3.469
56	3-NO <sub>2</sub> -4-Cl	0.74	-0.28	0.10	0.71	0.41	-0.15	1.0	6.30	5.80	6.355	4.20	5.02	4.537			
57	3-NO <sub>2</sub> -7-NH <sub>2</sub>	0.74	-0.28	0.54	0.71	0.00	0.00	0.0	5.77	5.73	5.998	5.17	4.91	5.140	4.77	4.22	4.376
58	3-NO <sub>2</sub> -4-NO <sub>2</sub>	0.74	-0.28	0.10	0.71	0.67	0.16	1.0	5.97	6.23	6.011	3.91	4.81	3.951			
59	3-NO <sub>2</sub> -4-F	0.74	-0.28	0.10	0.71	0.43	-0.34	1.0	6.00	5.84	6.013	5.17	5.23	5.140	4.88	4.26	4.502
60	3-NO <sub>2</sub> -4-F-7-CH <sub>3</sub>	0.74	-0.28	0.57	0.71	0.43	-0.34	0.0	6.50	6.42	6.554	5.80	5.47	5.626	5.50	4.47	4.826

<sup>a</sup>The data were taken from the literature (Ref. 34). <sup>b</sup>Anti-pentylenetetrazole effect. <sup>c</sup>Results by the multiregression analysis. <sup>d</sup>Results by the neural network.

表4-13. 平均值, 分散値, 標準偏差の比較

	anti-pent. <sup>a</sup>		anti-fight. <sup>b</sup>		clined scr. <sup>c</sup>	
	NN <sup>d</sup>	MR <sup>e</sup>	NN <sup>d</sup>	MR <sup>f</sup>	NN <sup>d</sup>	MR <sup>g</sup>
MD <sup>h</sup>	0.24	0.39	0.16	0.30	0.25	0.33
variance	0.11	0.37	0.04	0.15	0.09	0.17
SD <sup>i</sup>	0.33	0.61	0.21	0.39	0.31	0.41

<sup>a</sup>Anti-pentylentetrazole effect.

<sup>b</sup>Anti-fighting behavior.

<sup>c</sup>Clined screen test.

<sup>d</sup>Results by the neural network.

<sup>e, f, g</sup>Results calculated by eqs 33, 34, and 35 in Ref. 34, respectively.

<sup>h</sup>Mean deviation.

<sup>i</sup>Standard Deviation.



表4-14. ベンゾジアゼピンの Leave-n-0ut 法の分散値の平均

leave-n-out	$\sigma^2$	$\sigma_L^2$
1 <sup>b</sup>	0.280	0.081
2	0.460	0.084
5	0.539	0.076
10	0.913	0.077

<sup>a</sup>Applied to anti-pentylenetetrazole effct.

<sup>b</sup>Simple mean value of differences (observed value—calculated value).

## 第5章 ニューラルネットワーク法の活性相関研究の展開

ニューラルネットワーク法は従来法と比較して、非線形関係にある構造データと生理活性データ間の関係を付ける能力に優れているが、等級分類及びQSARで生理活性強度に対する入力データ種の役割を表現する定量的方法が無い点が不十分である。本章ではこの問題を検討する。

### 5.1 相関指数

本節ではニューラルネットワークへの線形性の導入方法及び出力強度に対する入力データ種の役割を知るための相関指数の求め方を提示し、その応用例を示す。

#### 5.1.1 ニューラルネットワークへの線形性動作の導入

4.2.2 節で議論したように、強い非線形性動作は不十分な測定データ数から未知データの活性強度を推測するときは必ずしも有効ではない。推測機能を強化するためには、ニューラルネットワークの動作に何らかの制約を導入し、この制約を測定データに含まれていると予想される関係から導かれるものと矛盾しない様にする。

QSARに適用出来る制約は「測定データが線形の関係に従う」ということである。ニューラルネットワークにこの制約を導入する現実的な手段は、ニューロンの動作関数をより線形動作に適した関数にすることである。

動作関数として、式1-4 の代わり、シグモイド関数と線形関数を一次結合させた関数式5-1 を用いる。

$$O_j = \beta / \{1 + \exp(-\alpha y_j)\} + (1 - \beta) y_j \equiv f(y_j),$$
$$y_j = (\sum w_{ij} x_i) - \theta_j \quad 5-1$$

この式で、 $\beta$  は 0~1 の値をとり、非線形性動作への線形性動作の混入の割合を示す。 $\beta$  値を変えることによって任意の線形性の制約をニューラルネットワーク

に課することができる。第2, 3層間の学習は式2-9~11に従い, 第1, 2層間の学習は第2層のニューロン値が0~1の場合は式2-9~11で行うことができる。0~1領域を外れる場合, 後に述べる疲労付きバックプロパゲーション学習にする必要がある。しかし $\beta$ 値が $\sim 0$ の場合, 0~1領域を外れることは殆ど無い。

動作関数の微分関数は,

$$\begin{aligned} f'(y_i) &= \alpha \beta g(y_i) \{1 - g(y_i)\} + 1 - \beta, \\ g(y_i) &= 1 / \{1 + \exp(-\alpha y_i)\} \end{aligned} \quad 5-2$$

である。式5-1, 5-2において,  $\varepsilon$ ,  $\alpha$ , 及び $\beta$ 値は層ごとに独立に設定することが可能である。

### 5. 1. 2 ニューラルネットワークにおける偏相関係数

複数の入力データと出力強度との間に何らかの関係があるとき, 入力データが出力に与える影響の程度, すなわち偏相関係数を求める方法を検討する。

入力, 教師データ, 及び出力ベクトルをそれぞれ,  $\{x_{k,i}\}$ ,  $\{t_i\}$ , 及び  $\{O_i\}$  で表す。ここで添字  $k$  及び  $i$  はデータの種類及びデータの番号を示す。偏相関係数は以下のような手順で求めることができる。

step 1:  $\{t_i\}$  及び データ種  $L$  を除いた入力データ  $\{x_{k(\neq L), i}\}$  を用いて学習させる。学習が成立したのち,  $\{x_{k(\neq L), i}\}$  をネットワークに与え, 出力ベクトル  $\{O_i\}$  を得る。それを基にベクトル差  $\{\Delta_i\} = (O_i - t_i)$  を求める。

step 2: 教師パターン  $\{x_{L,i}\}$  及びデータ種  $L$  についてのデータを除いた入力データ  $\{x_{k(\neq L), i}\}$  を用いて学習させ, 出力ベクトル  $\{O'_i\}$  を求め, ベクトル差  $\{d_i\} = (O'_i - x_{L,i})$  を得る。

step 3:  $\{\Delta_i\}$  と  $\{d_i\}$  との相関係数をつぎの式で求める。

$$r = \sum (\Delta_i - \bar{\Delta}) (d_i - \bar{d}) / \{ \sum (\Delta_i - \bar{\Delta})^2 \sum (d_i - \bar{d})^2 \}^{1/2} \quad 5-3$$

なお、 $\Delta$  及び  $d$  はそれぞれ  $\{\Delta_i\}$  及び  $\{d_i\}$  の平均値である。

ここで、重回帰分析法の偏相関係数とニューラルネットワーク法の偏相関係数の相違について考える。 $\alpha$  及び  $\beta$  値をそれぞれ 1 及び 0 に設定する。第 1 層には必ず 1 の値をとるニューロンを配置し、第 2 層のニューロン数を多くとった MR 型ニューラルネットワークの動作を考える。極限  $\beta^{(2)} \rightarrow 0$  ( $\beta^{(2)}$  は第 2 層のニューロンの  $\beta$  値を示す) では、3 層のネットワークは 2 層のネットワークの動作となり、このネットワークの学習は「2 乗誤差の和が極小になるような停留点」を求めることである。重回帰分析法と類似しているが、重要な違いがある。重回帰分析法は全入力データに対する 2 乗誤差の和の極小値を求めるのに対し、ニューラルネットワークは各々の入力データに対する 2 乗誤差の和を求めている。その結果、停留点の相違は入力データ数が多くなるに従って大きくなる。

$\beta^{(2)} \rightarrow 1$  の極限ではネットワークの動作は非線形になり、入力データの差が少なくとも、それを強調して教師データに対応づける性質が強まる。この能力が大きいほど分類や入力データの再現性に優れるが、ベクトル  $\{d_i\}$  が次第に 0 ベクトルに近づくため step 3 における偏相関係数を求める際に困難が生じる。すなわち、偏相関係数 ( $r$ ) はネットワークの学習の仕方に依存する。 $r$  を求めるには 2 つの独立したネットワークを学習させる必要があるため、両方に対し等価な学習を施さないかぎり、 $r$  値は大きく変動する。この性質は適切ではない。以上の方法を方法 1 とする。

### 5. 1. 3 ニューラルネットワークにおける相関指数

実用に供せられる偏相関係数を求めるため、一回の学習から係数を求める必要がある。一手段として、入力データの微小変化が出力強度に与える変化を求める方法がある。すなわち、入力データの微小変化  $\delta x_L$  は第 2 層のニューロン値に

$\delta y_j$  の変化を与える。

$$\delta y_j = w_{Lj} \delta x_L \quad 5-4$$

$L$  は微小変化  $\delta$  が起こったデータを示す。ニューロンの出力値の変化は、

$$\delta O_j = \beta g'(y_j) \delta y_j + (1 - \beta) \delta y_j \quad 5-5$$

である。ここで、 $g'$  はシグモイド関数の微分関数 (式2-7) である。微分関数を代入して、

$$\delta O_j = \beta \alpha O(y_j) \{1 - O(y_j)\} \delta y_j + (1 - \beta) \delta y_j \quad 5-6$$

を得る。したがって、入力データと第2層のニューロン  $j$  の出力強度の微小変化の比は、

$$\delta O_j / \delta x_L = [\beta \alpha O(y_j) \{1 - O(y_j)\} + (1 - \beta)] w_{Lj} \quad 5-7$$

で与えられる。同様に、入力データの微小変化とネットワークの出力の微小変化との比は、

$$\begin{aligned} \delta O_k / \delta x_L = & \sum [\beta^{(3)} \alpha^{(3)} O(z_k) \{1 - O(z_k)\} \\ & + (1 - \beta^{(3)})] w_{jk}^{(2,3)} \\ & \times [\beta^{(2)} \alpha^{(2)} O(y_j) \{1 - O(y_j)\} + (1 - \beta^{(2)})] w_{Lj}^{(1,2)} \end{aligned} \quad 5-8$$

となる。ここで、 $O(z_k)$  は入力データに微小変化がないときの出力であり、 $O(y_j)$  は第2層のニューロン  $j$  の出力である。 $\beta$ 、 $\alpha$ 、及び  $w$  にある添字は層の番号を示す。入出力が非線形の関係にあるときは、式5-8 で与えられる  $\delta O_k / \delta x_L$  は  $x_L$  の値によって異なる。重回帰分析の結果と比較するときは、入出力の線形関係が想定出来る場合に限り、 $\delta O_k$  値を各入力データについて平均化する。これを偏相関指数と呼ぶことにする。

$$(\delta O_k / \delta x_L) = N^{-1} \sum_p (\delta O_k / \delta x_L)_p \quad 5-9$$

ここで、 $N$ は入力データの数を示す。 $\beta \rightarrow 1$ のとき、偏相関指数 $(\delta O_k / \delta x_L)$ は、出力値が0または1付近で、0に近づく。この性質はニューロンの動作関数であるシグモイド関数に由来する。QSARに適用するニューラルネットワークでは、生理活性の最大強度を1に設定する。それゆえ上記の性質は適切ではない。この問題は、ニューロンの動作関数として、関数値0あるいは1において微分値が0とならないような関数を用いることによって軽減できる。そのような関数の一つに、次式で定義される不飽和シグモイド関数がある。

$$f(y) = C / \{a + \exp(-\alpha y)\} \quad 5-10$$

パラメータ $C$ 及び $a$ を $f(1) = 1$ 及び $f(0) = 0.5$ となるように決定する。例えば、 $\alpha = 1$ 、 $a = 0.2642411$ 、及び $C = 0.6321206$ に設定すると要求されている条件を満たすことができる。

このような関数を用いバックプロパゲーション学習を行うと、ニューロン値が定義域(0~1)から外れる。この問題を取り扱うために、ニューロンの疲労という概念を導入する。すなわち、学習の過程でニューロン値が1を越えたとき、そのニューロンと結合する全てのニューロンとの間 $w_{ij}$ の学習を一時停止し、一定時間を経過したのち、ニューロン値が1となるように $w_{ij}$ に定数を乗じ、学習に復帰させる。この操作はニューロンが疲れたとき「休ませる」ことであり、学習過程でニューロンの疲労を取り入れたことに相当する。式5-10に示した関数の微分関数は、

$$f'(y) = C \alpha h(y) \{1 - a h(y)\} \quad 5-11$$

$$f(y) = C / \{a + \exp(-\alpha y)\} \equiv C h(y) \quad 5-12$$

である。

$w_{ij}$ の学習の一時停止期間は、バックプロパゲーションの操作回数で示すと、9~20回が適切であった。この学習法は制約下の変分操作であるため、学習の完了が保証出来ない。しかし、実際QSARに適用したところでは、第2層のニューロン数を通常のバックプロパゲーション時の2~3倍以上に設定すると学習未完の問題は発生しなかった。不飽和シグモイド関数はシグモイド関数と線形関数の中間のfitting能力を持っているので、第2層のニューロン数を多めに取る必要がある。この方法を方法2とする。<sup>39)</sup>

#### 5. 1. 4 カルボキノン誘導体のQSARへの応用

〔方法1の適用〕 4.2.1節において用いたQSARデータに適用した。入力データの種類数は6であるので、バイアスを加えて、ネットワーク構造を  $N(7,6,1)$  とした。パラメータ $\alpha$ 、及び $\theta$ 値をそれぞれ、1.0、及び0.0に設定し、 $\epsilon$ 値は、第1、2層間には0.06、また第2、3層間には0.03とした。結果を表5-1に示す。

はじめに $\alpha$ と $\beta$ の組み合わせと決定係数との関係について考察した。 $\beta=1$ 、 $\alpha=5$ としてネットワークの非線形動作を強制的に強めると決定係数は大きく、計算値と観測値との一致が良くなる。それに対し、線形性動作を強調すると決定係数が下がり、重回帰分析法とのそれに近くなる( $\alpha=1$ 、 $\beta=0.2$ の場合を参照)。これらの結果は5.1.1節の議論で予想された通りである。一方、ニューロンの動作関数が変わると、 $\pi_2$ とFに関する相関指数が大きく変化する。表5-2に入力データ間の相関係数を示す。この表から $\pi_2$ と $\pi_{1,2}$ 、FとRに強い相関関係がある。したがって、動作関数及び学習の方法が変わることによって、 $\pi_2$ 、 $\pi_{1,2}$ 、F、Rの内少なくとも2つの変数に大きな相関指数値の変動が計算されることが理解出来る。

表5-1. カルボキノン誘導体における種々の $\alpha$ ,  $\beta$ 値による偏相関係数

$\alpha$	5	2	1	1	1	MR <sup>a</sup>
$\beta$	1	1	1	0.5	0.2	
MR <sub>1,2</sub>	-0.01	-0.07	-0.09	-0.07	-0.11	-0.04
$\pi_{1,2}$	-0.35	-0.33	-0.22	-0.15	-0.27	-0.28
$\pi_2$	-0.18	-0.17	-0.35	-0.48	-0.30	-0.59
MR <sub>1</sub>	-0.25	-0.38	-0.37	-0.42	-0.40	-0.29
F	0.06	-0.11	-0.43	-0.50	-0.45	-0.30
R	-0.38	-0.41	-0.45	-0.42	-0.48	-0.27
DC <sup>b</sup>	0.924	0.889	0.883	0.851	0.850	0.853

<sup>a</sup>Results by multiregression analysis.

<sup>b</sup>Decision coefficient.



表5-2. 重回帰分析によって得られた変数間の偏相関係数

	$MR_{1,2}$	$\pi_{1,2}$	$\pi_2$	$MR_1$	F
$\pi_{1,2}$	-0.01				
$\pi_2$	-0.02	0.72			
$MR_1$	0.48	0.38	-0.55		
F	0.04	-0.09	-0.27	-0.16	
R	-0.28	0.07	-0.33	-0.30	-0.72

〔方法2の適用〕 同じデータを用いて方法2による相関指数を求めた。ネットワーク構造は  $N(7,18,1)$  とした。動作関数は標準的な不飽和シグモイド関数、

$$f(y) = 0.6321206 / \{0.2642411 + \exp(-y)\} \quad 5-13$$

を用い、 $\epsilon$  及び  $\theta$  はそれぞれ 0.06 及び 0 に設定し、不活性化のサイクルを 9 とした。出力層のニューロンの動作関数を  $f(y) = y$  とし、 $\epsilon$  及び  $\theta$  値はそれぞれ 0.03 及び 0 とした。入力データは 0.1~1.0 間にスケールし、学習サイクル数を 30000 とした。結果を表5-3 に、標準偏回帰係数 (SPRC) 及び偏回帰係数 (PRC) とともに示す。

MED での決定係数の低下は、不飽和シグモイド関数がシグモイド関数の一部分であるため、非線形の fitting 能力が通常のシグモイド関数に比べて低下するためである。相関指数 ( $\delta O_j / \delta x_L$ ) は相対的な指標であり、値及びその符号が重要である。これらを重回帰分析法の相関係数と比較すると、値の傾向が良く一致する。すなわち、生物活性にもっとも強く反映している入力データ種は両方法で一致する。

方法1と方法2を比較すると次のようなことがいえる。(1) 方法2によって相関指数を求める方法は方法1に比べて 1/20 程の計算量で済み、しかも2つのネットワークを学習する必要がない。(2) 方法2では相関指数を求めたネットワークを用いて決定係数を求めることができる。

### 5. 1. 5 ベンゾジアゼピン類のQSARへの適用

方法2をベンゾジアゼピン類のQSARへの適用した。ネットワーク構造を  $N(8,21,1)$  とし、学習サイクル数を 20000 に設定した。結果を表5-4 に示す。

すでに述べたようにこの生物活性データには多くの誤差及び非線形関係を含む。これは重回帰分析法による決定係数 (0.388 ~ 0.493) が低いことより判定できる。一方、ニューラルネットワーク法を用いると非線形 fitting が働き決定係数

表5-3. カルボキノン誘導体の相関指数 ( $\delta O_k / \delta x_L$ )

Variable	連続投与				一回投与			
	MED <sup>a</sup>		OD <sup>b</sup>		MED <sup>a</sup>		OD <sup>b</sup>	
	N.N <sup>c</sup>	SPRC <sup>d</sup>	PRC <sup>e</sup>	N.N <sup>c</sup>	SPRC <sup>d</sup>	PRC <sup>e</sup>	N.N <sup>c</sup>	SPRC <sup>d</sup>
MR <sub>12</sub>	-6	-0.04	-0.02	-8	-0.13	-0.06	2	0.07
$\pi_{12}$	-16	-0.18	-0.09	-24	-0.38	-0.17	-22	-0.46
$\pi_2$	-19	-0.59	-0.42	-6	-0.29	-0.17	-10	-0.35
MR <sub>1</sub>	-11	-0.29	-0.27	-9	-0.19	-0.15	-10	-0.19
F	-16	-0.30	-1.77	-23	-0.43	-2.04	-35	-0.70
R	-8	-0.27	-0.75	-17	-0.43	-1.00	-20	-0.58
DC <sup>f</sup>	0.868	0.853	0.858	0.822	0.854	0.839	0.893	0.844

<sup>a</sup>Minimum effective dose.

<sup>b</sup>Optimal dose.

<sup>c</sup>Results by the neural network.

<sup>d</sup>Standard partial regression coefficient.

<sup>e</sup>Partial regression coefficient.

<sup>f</sup>Decision coefficient.

表5-4. ベンゾジアゼピン誘導体の相関指数 ( $\delta O_k / \delta x_L$ )

Variable	anti-pent. effect <sup>a</sup>		anti-fighting effect		aligned screen test				
	N.N <sup>b</sup>	SPRC <sup>c</sup>	PRC <sup>d</sup>	N.N <sup>b</sup>	SPRC <sup>c</sup>	PRC <sup>d</sup>			
MR-3	11	-0.12	-0.22	-9	0.01	0.02	5	0.04	0.08
$\pi$ -3	15	0.10	0.13	29	0.02	0.02	42	0.17	0.18
MR-7	-15	-0.27	-0.20	-20	-0.48	-0.26	-42	-0.63	-0.35
$\sigma$ -3	17	0.51	1.85	32	0.46	1.22	28	0.45	1.26
F-4	17	0.29	1.13	16	0.13	0.36	-4	-0.05	-0.14
R-4	-12	-0.02	-0.12	-2	-0.16	-0.58	0	-0.09	-0.33
I-1	-5	-0.21	-0.34	-9	-0.41	-0.47	-9	-0.33	-0.40
DC <sup>b</sup>	0.787	0.493	0.828	0.443	0.809	0.388			

<sup>a</sup>Anti-pentylene-tetrazole effect.

<sup>b</sup>Results by the neural network.

<sup>c</sup>Standard partial regression coefficient.

<sup>d</sup>Partial regression coefficient.

<sup>e</sup>Decision coefficient.

が大幅に改善される。

相関指数は重回帰分析法の SPRC 及び PRC 値に似た傾向を示すが、前例のカルボキノン類の QSAR の場合より、類似性が少ない。これはニューラルネットワークでは非線形動作が強く働くためである。<sup>39)</sup>

## 5. 2 再構築学習法

バックプロパゲーションアルゴリズムはニューラルネットワークの中に入力データから教師データを作り出す「変換行列を生成する操作」である。それゆえ、逆の「変換行列を消去する操作」を考えることが出来る。多くの消去操作の中から、ニューロン間の結合荷重が最も学習容易な方向に消去する操作が存在する。この操作とバックプロパゲーション操作を交互に行い、教師データとニューラルネットワークの出力の2乗誤差が一定になるように保つならば、それはニューロン間の結合荷重の中で上記の変換を実現する必要最小限の結合荷重を強調する状態になっている可能性がある。この淘汰状態を学習の一過程でとることを再構築学習法<sup>38)</sup>という。

### 5. 2. 1 再構築学習法の方法

再構築学習法は学習状態と淘汰状態から成り、学習状態はバックプロパゲーション操作から構成される。淘汰状態は学習と消去過程から構成される。学習過程はバックプロパゲーション操作を行い、消去過程は式5-14 に従い、ニューロン間の結合荷重の絶対値を0に近づけることである。それは結合荷重が最も学習容易な方向に消去する操作の一つである。

$$w_{ij} = w_{ij} - \text{sgn}(w_{ij}) \zeta \{1 - D(w_{ij})\} \quad 5-14$$

ここで、Dは $|w_{ij}| < \zeta$  のとき 1,  $|w_{ij}| > \zeta$  のとき 0を与える関数

であり、 $w_{ij}$  を急激に変化させないように、 $\xi$  を  $\epsilon$  の約1/10 程度に設定する。

{ } 内は  $w_{ij}$  の絶対値が非常に0に近い時、消去によってその符号が変化しないように追加された項である。

ある層の  $j, k$  番目のニューロンが同じ種類の情報に関与しており、これらのニューロンに一つ前の層のニューロン  $i$  から  $w_{ij}, w_{ik}$  という結合を介して情報が伝播しているとする。ただし  $w_{ij} > w_{ik}, \text{sgn}(w_{ij}) = \text{sgn}(w_{ik})$ 。

バックプロパゲーション操作によって  $w_{ij}, w_{ik}$  が  $\delta w_{ij}, \delta w_{ik}$  だけ変化する。すると式2-9によって、 $\delta w_{ij} / \delta w_{ik} = w_{ij} / w_{ik}$  である。式5-14から本質的でない { } 内の項を除いて、 $\delta w_{ij} = -\text{sgn}(w_{ij}) \xi$ 。  $\xi \ll \delta$  の下で、

$$(\delta w_{ij} - \xi) / (\delta w_{ik} - \xi) > \delta w_{ij} / \delta w_{ik}.$$

すなわち消去過程によって、 $w_{ij}, w_{ik}$  の比は拡大する。これが誤差関数  $E$  一定のままで繰り返されると、 $|w_{ij}| \gg |w_{ik}|$  となって、「同種の情報を生成する複数のニューロンが淘汰されて1個のニューロンに集約」される。

淘汰状態は以上の機能を有するが、これだけでは誤差関数  $E$  を0に接近させる機能が無く学習完了に至らないため、学習状態を時々取る必要がある。ここでは学習、淘汰状態を分離して示したが、両状態を混合して非常にゆっくり  $E$  を0に接近させる事も可能である。しかし学習効率を考えると  $E \sim 0.1$  までは学習状態だけを取らせた方が良いようである。

再構築学習法で学習すると情報の通路関係にないニューロンの孤立が起こり、入出力間の情報の流れが明確になる。したがって、結合を辿れば特定の入力データの出力への役割を解析することができる。

## 5. 2. 2 ノルボルネンにおける $^{13}\text{C}$ -NMRケミカルシフトとコンホメーションの相関関係の問題への適用

ノルボルネンにおける $^{13}\text{C}$ -NMRケミカルシフトとコンホメーションの相関

関係の問題に対し再構築学習法を適用する。ネットワーク構造は 4.1.1 節と同様にする。化合物 1~25 を学習用に用い、29~38 を未知化合物として推測する。

再構築学習法を学習と淘汰の 2 状態から構成し、 $E \sim 0.1$  まで学習状態のみを取り、それ以下では 2 状態を交互に取るようにした。淘汰状態は 1 回の学習過程に対し 1 回の消去過程から構成した。このとき対応する学習状態は 10~50 回のバックプロパゲーション操作から構成した。この回数は  $E$  値が前回の  $E$  値を超えないようにプログラム内部で決定した。

表 5-5 には学習状態のみの通常学習を行ったときの結合荷重行列と学習化合物と未知化合物に対する分類結果を示し、表 5-6 には上記の再構築学習法での結果を示す。ただし、結合の強度を比較するため、結合の最大値を 0.999 にスケールした（スケール因子は各表の脚注に示した）。

4.1.1 節で述べたように、通常学習法でのニューラルネットワーク法の結果は線形学習機械法及びクラスター分析法<sup>21)</sup>より優れている。出力層に 2 個のニューロンを用いた今回の場合も分類、推測能力は 4.1.1 節の場合と同じである。各層間の結合荷重行列をみると、通常学習法ではすべてのニューロンが結合し、情報が分散している。一方表 5-5 に示すように再構築学習を行うと、ネットワークの分類、推測能力は通常の学習の場合と同等であるが、結合荷重行列に情報の局在化が起こり、情報の進行経路は単純化される。

出力層の 1 番目のニューロンの発火は *exo* を、2 番目は *endo* に対応し、第 2 層のニューロン 6 と 8 が出力層のニューロンと  $(a, -a)$  及び  $(-b, b)$  の結合でコンホメーションを決定している。また、第 2 層のニューロン 6 と 8 は入力層のニューロン 6 と 7 と最も強く結合している。入力層のニューロンの番号はノルボルネン骨格の炭素原子の番号に対応させてあるので、以上の結果は  $C_6$  と  $C_7$  のケミカルシフトに関する情報が、ノルボルネン誘導体の *endo/exo* 決定に中心的役割を演じていることを示す。これは  $C_6$  と  $C_7$  が置換基の近くにあり、

表5-5. ノルボルネン誘導体<sup>a</sup>の<sup>13</sup>C-NMRケミカルシフト計算時の結合荷重行列と出力値

A <sup>b</sup>														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	-0.022	0.133	-0.188	-0.147	0.047	0.063	-0.009	0.112	-0.044	-0.214	-0.117	-0.112	0.092	-0.221
2	0.219	0.203	0.240	-0.040	0.108	0.265	-0.355	-0.483	0.333	-0.054	-0.099	0.230	-0.358	0.011
3	0.321	-0.043	-0.013	-0.173	-0.141	0.083	-0.143	-0.186	0.156	-0.035	-0.029	0.097	-0.182	0.128
4	-0.233	-0.149	-0.273	-0.192	0.205	-0.559	0.489	0.451	-0.339	-0.023	0.043	-0.350	0.445	-0.242
5	-0.406	-0.095	-0.097	0.120	0.093	-0.261	0.367	0.324	-0.429	-0.149	0.318	-0.288	0.381	0.067
6	0.449	0.068	0.421	0.006	-0.095	0.664	-0.760	-0.747	0.524	0.180	-0.251	0.358	-0.535	0.276
7	-0.773	-0.381	-0.495	-0.086	-0.038	-0.716	0.905	0.999	-0.645	-0.235	0.583	-0.278	0.615	-0.404
-----														
B <sup>c</sup>														
1	2	C <sup>d</sup>			D <sup>e</sup>			E <sup>f</sup>			F <sup>g</sup>			
		1	2	3	1	2	3	1	2	3	1	2	3	
1	0.751	-0.732	1	0.913	0.084	14	0.014	0.985	26	0.703	0.293	26	0.703	
2	0.221	-0.374	2	0.989	0.009	15	0.015	0.983	27	0.079	0.924	27	0.079	
3	0.624	-0.387	3	0.994	0.005	16	0.088	0.900	28	0.906	0.091	28	0.906	
4	0.149	0.038	4	0.844	0.155	17	0.020	0.979	29	0.025	0.974	29	0.025	
5	-0.217	-0.078	5	0.955	0.049	18	0.017	0.984	30	0.981	0.018	30	0.981	
6	0.747	-0.876	6	0.961	0.039	19	0.025	0.975	31	0.708	0.293 <sup>h</sup>	31	0.708	
7	-0.913	0.970	7	0.948	0.051	20	0.014	0.986	32	0.953	0.045	32	0.953	
8	-0.998	0.956	8	0.964	0.034	21	0.220	0.788	33	0.009	0.990	33	0.009	
9	0.692	-0.746	9	0.926	0.073	22	0.015	0.984	34	0.974	0.029	34	0.974	
10	0.188	-0.237	10	0.968	0.030	23	0.067	0.926	35	0.063	0.941	35	0.063	
11	-0.437	0.453	11	0.970	0.028	24	0.021	0.979	36	0.008	0.992	36	0.008	
12	0.431	-0.445	12	0.974	0.026	25	0.019	0.981	37	0.825	0.177	37	0.825	
13	-0.655	0.703	13	0.841	0.161				38	0.103	0.898	38	0.103	
14	0.467	-0.219												
-----														

<sup>a</sup>The sum of the squared errors (E) < 0.110.

<sup>b</sup>Weight matrix between the first (column) and second (row) layers (scale factor = 0.4176).

<sup>c</sup>Weight matrix between the second (column) and third (row) layers (scale factor = 0.390). <sup>d</sup>Outputs for the training data. <sup>e</sup>Outputs for the untrained data. <sup>f</sup>Error.



表5-6. 再構築学習法<sup>a)</sup>による結合荷重行列と出力値

A <sup>b</sup>															
1	2	3	4	5	6	7	8	9	10	11	12	13	14		
1	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.000		
2	0.000	0.000	0.000	0.000	0.000	0.594	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
3	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
4	0.000	0.000	0.000	0.000	0.000	-0.515	0.000	0.006	0.000	0.000	0.000	0.000	0.000		
5	0.000	0.000	0.000	0.000	0.000	-0.626	0.000	0.003	0.000	0.000	0.000	0.000	0.000		
6	0.000	0.000	0.000	0.000	0.000	0.994	0.000	-0.122	0.000	0.000	0.000	0.000	0.000		
7	0.000	0.000	0.000	0.000	0.000	-0.998	0.000	0.744	0.000	0.000	0.000	0.000	0.000		

B <sup>c</sup>				C <sup>d</sup>				D <sup>e</sup>			
1	2	cmpd	exo	cmpd	exo	endo	exo	endo	cmpd	exo	endo
1	-0.001	0.001	1	0.899	0.101	14	0.050	0.950	26	0.707	0.293
2	-0.001	0.001	2	0.973	0.027	15	0.051	0.949	27	0.115	0.885
3	-0.001	0.001	3	0.978	0.022	16	0.136	0.864	28	0.904	0.096
4	-0.001	0.001	4	0.855	0.145	17	0.063	0.937	29	0.070	0.930
5	-0.001	0.001	5	0.939	0.061	18	0.049	0.951	30	0.966	0.034
6	0.998	-0.998	6	0.955	0.045	19	0.072	0.928	31	0.629	0.371 <sup>f</sup>
7	0.001	0.001	7	0.931	0.069	20	0.046	0.954	32	0.941	0.059
8	-0.577	0.577	8	0.947	0.053	21	0.226	0.774	33	0.041	0.959
9	-0.001	0.001	9	0.914	0.086	22	0.050	0.950	34	0.958	0.042
10	-0.001	0.001	10	0.935	0.065	23	0.115	0.885	35	0.100	0.900
11	-0.001	0.001	11	0.959	0.041	24	0.059	0.941	36	0.038	0.962
12	-0.001	0.001	12	0.962	0.038	25	0.042	0.958	37	0.738	0.262
13	-0.001	0.001	13	0.759	0.241				38	0.093	0.907
14	-0.001	0.001									

<sup>a)</sup>The sum of the squared errors (E) < 0.153. <sup>b)</sup>The weight matrix between the first (column) and second (row) layers (scale factor=0.3227). <sup>c)</sup>The weight matrix between the second (column) and third (row) layers (scale factor = 0.1681). <sup>d)</sup>Outputs for the trained data. <sup>e)</sup>Outputs for the untrained data. <sup>f)</sup>Error.

置換基効果が  $C_6$  と  $C_7$  では逆転するという化学的常識に一致する。第2層のニューロン 8 は第1層のニューロン 6 及び 7 と結合している。これは第2層のニューロン 6 に含まれる endo/exo に関する情報をニューロン 8 で修正することを示す。この系では、再構築学習法は必要な最小限のニューロン数を与えているようである。第2層のニューロン数を14 としたが、2個で十分であり、第1層のニューロン1及び 3 に対応する情報は endo/exo の分離には関係ない。構造活性相関の研究では説明変数の数を最小限にすることが要請されている。再構築学習法はこのような要求を満たす一手法と考えられる。

### 5. 2. 3 マイトマイシンの等級分類への応用

表5-7 に再構築学習法による結合荷重行列及び分類結果を示す。用いた入力データ、ネットワーク構造は 4.1.2 節で用いたものと同じである。

表5-7A の5及び6番目の行に注目する。それらは第2層のニューロン 5 及び 10 と強く結合している。表5-7B から、第2層のニューロン 10 と 5 は出力層の強度 3+ と 2+ に対応するニューロンに強く正結合し、+/- と - に対応するニューロンには結合しない。第2層には出力層の 3+ ニューロンを強く抑制するニューロンはないが、2+ ニューロンを抑制するニューロンは 2 と 10 である。また、- ニューロンを抑制するためにはニューロン 5 を発火しなくてはならない。これはニューロン 2+ 及び + の活性化に相当する。したがって、出力層の 3+ ニューロンを発火させる因子は、置換基 Z の大きな Taft の立体定数 ( $E_{S-Z}$ ) 及び置換基 X の低い  $F_x$  値及び  $V_{w-x}$  値である。以上の結果と入力データ種  $\sigma_{m-x}$  の寄与がほとんどないことは ALS 法による結果を支持するが、ダミー定数、 $Y_{OMe}$ 、及び  $Y_{OH}$  が生物活性に寄与しないという結果は、ALS法と矛盾する。一方 2+ ニューロンは第2層のニューロン 5 を通して発火するが、そのニューロンは  $E_{S-Z}$  及び  $Y_{OMe}$  によって規定される。置換基 X の存在は 2+ ニューロンをわづ

Table 5-7. マイトマイシンの再構築学習法の結合荷重行列と出力値

A <sup>b</sup>												
1	2	3	4	5	6	7	8	9	10	11	12	
1	0.000	0.000	0.000	0.001	-0.064	0.001	0.000	0.001	0.000	-0.355	0.000	0.000
2	0.000	0.682	0.000	0.001	-0.044	0.000	-0.073	0.000	0.000	0.000	0.000	0.000
3	0.000	0.515	0.000	0.000	0.000	-0.039	0.000	0.000	0.000	-0.211	0.000	0.000
4	0.000	0.001	0.000	-0.001	0.585	-0.001	0.010	0.000	-0.001	-0.032	0.000	0.000
5	0.000	0.000	0.000	0.005	0.113	0.075	-0.201	0.000	0.000	-0.121	0.000	0.000
6	0.000	-0.698	0.000	0.000	0.999	0.000	0.000	0.000	0.000	0.398	0.000	0.000
7	0.000	0.000	0.000	0.064	-0.682	0.065	0.174	0.125	0.000	-0.110	0.000	0.000

B <sup>c</sup>												
1	2	3	4	5	C <sup>d</sup>							
					cmpd	3+	2+	+	+/-	-	obs	
1	-0.001	0.001	-0.001	0.000	0.000	1	0.893	0.075	0.016	0.001	0.012	3+
2	-0.001	-0.827	0.813	0.333	0.001	2	0.942	0.037	0.018	0.001	0.014	3+
3	-0.001	0.000	-0.164	-0.001	0.000	3	0.039	0.947	0.030	0.001	0.012	2+
4	-0.001	0.001	-0.378	0.000	0.001	4	0.032	0.951	0.034	0.001	0.012	2+
5	0.000	0.529	0.850	-0.001	-0.937	5	0.013	0.900	0.157	0.002	0.013	2+
6	0.000	0.000	-0.812	-0.001	0.000	6	0.011	0.916	0.000	0.080	0.013	2+
7	-0.007	-0.001	0.000	-0.760	0.562	7	0.065	0.938	0.019	0.001	0.012	2+
8	-0.529	0.001	-0.262	-0.001	0.000	8,9	0.007	0.038	0.760	0.012	0.110	+
9	-0.001	0.001	0.000	-0.001	0.001	10	0.041	0.026	0.945	0.011	0.012	+
10	0.892	-0.999	0.000	-0.001	0.000	11	0.010	0.069	0.961	0.027	0.007	+
11	-0.001	0.000	0.000	0.000	0.000	12	0.007	0.042	0.988	0.042	0.007	+
12	0.000	0.001	-0.001	0.000	0.000	13	0.006	0.001	0.067	0.893	0.076	+/-
						14	0.087	0.095	0.000	0.001	0.948	-
						15	0.006	0.002	0.159	0.017	0.884	-
						16	0.006	0.000	0.023	0.054	0.978	-

<sup>a</sup>The sum of the errors (E) < 0.057. <sup>b</sup>The weight matrix between the first (column) and second (row) layers (scale factor = 0.0631). <sup>c</sup>The weight matrix between the second (column) and third (row) layers (scale factor = 0.0951). <sup>d</sup>Outputs by the reconstruction-learning method.

かに抑制するが、この結果は ALS 法の結論とは矛盾しない。ALS 法は X と Y の変数を、X が小さくなるにつれて Y が大きくなるように、一緒に認識するが、ニューラルネットワークはそれらを分離して認識している。しかし、活性値に対する寄与が OH より OMe が大きいことは両方法とも一致する。

第2層のニューロン 2 及び 10 は 2+ ニューロンを抑制する。ニューロン 10 と 2+ ニューロンの結合及びニューロン 10 とニューロン +3 の結合の符号は反対であることより、ネットワークはそれらを別々に認識することを示す。入力層から出力層のニューロン 5 への経路はニューロン 2 へのそれと符号を異にする。すなわち、変数  $E_{s-z}$  の寄与は小さく、 $\sigma_{m-x}$  及び  $V_{w-x}$  が大きい。この結果は ALS法のそれと一致する。

ALS法との関連をさらに詳しくみるため、等級 n に対し n 個の出力層のニューロンを発火させる教師パターンをネットワークに提示した。例えば、等級 ±, +, 等に対し、教師パターン(1,1,0,0,0), (1,1,1,0,0), 等を提示するのである。(この形の教師パターンを ALS型教師パターンと呼ぶ)。マイトマイシンのデータに ALS型教師パターンを与えて再構築学習させた結果を表5-8 に示す。

生物活性に関しては第2層の2, 3, 及び5番目のニューロンが重要な役割を果たすことが表からわかる。表5-9 に構造パラメータとそれらのニューロンとの結合を抽出して示す。ただし、抑制的ニューロンに結合した係数には -1 をかけてある。一方、ALS法による認識関数は次のように報告されている。<sup>10)</sup>

$$L = -4.10 \sigma_{m-x} - 1.76V_{w-x} + 2.49Y_{OMe} + 2.29Y_{OH} + 0.79$$

5-15

$$L = -4.33F_x - 2.48V_{w-x} + 2.47Y_{OMe} + 2.28Y_{OH} + 0.78E_{s-z} + 1.42$$

5-16

これの式と表5-9 比較すると、式5-14 は path 2 に対応し、式5-16 は path 3 と 5 の和に対応する。

表 5-8. マイトマイシン誘導体<sup>a</sup>におけるA L S型の学習パターン使用時の再構築学習法の結合荷重行列と出力値

A <sup>b</sup>												
1	2	3	4	5	6	7	8	9	10	11	12	
1	0.000	0.043	-0.065	0.000	0.499	-0.001	0.000	0.001	-0.001	0.000	0.001	0.000
2	0.000	0.479	-0.027	0.001	0.001	-0.001	0.000	0.000	-0.002	-0.001	0.000	0.000
3	0.000	0.403	-0.001	0.000	0.155	0.000	-0.001	0.001	0.000	0.000	-0.001	0.000
4	0.000	-0.119	0.633	0.001	0.002	0.000	0.000	0.000	0.060	-0.001	0.000	0.001
5	0.001	-0.165	0.268	-0.002	0.003	0.001	0.001	0.001	-0.005	0.002	0.000	0.001
6	0.000	-0.293	0.999	0.001	-0.389	0.001	0.001	0.001	0.000	0.001	0.001	0.001
7	0.001	0.002	-0.816	0.000	0.051	0.000	0.002	0.000	0.149	0.000	0.000	0.000

B <sup>c</sup>												
1	2	3	4	5	cmpd	-	+/-	+	2+	3+	obs	
1	0.001	0.000	0.000	0.000	1	0.975	0.999	1.000	0.992	0.795	3+	
2	0.002	0.000	-0.659	-0.866	2	0.973	0.999	1.000	0.987	0.908	3+	
3	0.053	0.999	0.867	0.505	3	0.981	0.998	0.999	0.960	0.104	2+	
4	0.002	-0.001	0.000	0.000	4	0.981	0.998	0.999	0.954	0.085	2+	
5	0.068	-0.002	-0.001	0.000	5	0.982	0.997	0.998	0.872	0.024	2+	
6	0.001	-0.001	-0.001	0.000	6	0.972	0.966	0.993	0.883	0.050	2+	
7	0.001	-0.001	0.000	0.001	7	0.981	0.999	0.999	0.930	0.098	2+	
8	0.063	-0.002	0.000	0.001	8,9	0.980	0.806	0.779	0.139	0.001	+	
9	0.224	-0.437	-0.001	0.000	10	0.980	0.999	0.966	0.080	0.047	+	
10	0.002	-0.002	-0.001	0.000	11	0.985	0.999	0.958	0.063	0.002	+	
11	0.001	0.000	0.000	0.001	12	0.986	0.998	0.893	0.023	0.000	+	
12	0.002	-0.001	0.001	0.000	13	0.977	0.852	0.116	0.002	0.000	+/-	
					14	0.963	0.084	0.092	0.021	0.004	-	
					15	0.977	0.194	0.142	0.017	0.000	-	
					16	0.971	0.043	0.001	0.000	0.000	-	

<sup>a</sup>The maximum sum of errors (E) < 0.099. <sup>b</sup>The weight matrix between the first (column) and second (row) layers (scale factor = 0.1066). <sup>c</sup>The weight matrix between the second (column) and third (row) layers (scale factor = 0.0894). <sup>d</sup>Outputs.

表5-9. マイトマイシンの生理活性発現に重要なパスと入力パラメータの寄与

path	$F_x$	$\sigma_{m-x}$	$V_{w-x}$	$Y_{ome}$	$Y_{oh}$	$E_{s-z}$
2	-0.04	-0.48	-0.40	0.12	0.17	0.29
3	-0.07	-0.03	0.00	0.63	0.27	0.99
5	-0.50	0.00	-0.16	0.00	0.00	0.39

#### 5. 2. 4 アリルアクリロイルピペラジンの等級分類への応用

標記問題について、ALS型教師パターンを与え学習させ、ALS法との結果と比較した。結果を表5-10に示す。分類は完全に達成できた。第2層のニューロン 1, 2, 4, 6, 及び 14 が重要な役割を演じる。それらのうち、出力層の最強活性に対応するニューロン 4 に正結合しているのは第2層のニューロン 4 と 14 であり、ニューロン 1 と 2 とは負結合である。これはニューロン 4 の発火因子はニューロン 3 のそれと異なることを示す。

表5-11に構造パラメータとの結合値を示す。ただし、抑制的ニューロンに結合した係数には -1 をかけている。一方、ALS法による認識関数は次のように報告されている。

$$L = -1.018\Delta RI - 2.357\Sigma\sigma - 1.838I(2-OR) + 0.769$$

5-17

$$L = -1.259\Sigma\pi - 2.166\Sigma\sigma - 1.777I(2-OR) + 1.169I(R1) + 0.358$$

5-18

しかし、これらの式による認識率は悪く、それぞれ、76% 及び 65% である。ニューラルネットワークは、入力データ種  $\Delta RI$ ,  $\Sigma\sigma$ , 及び  $\Sigma\pi$  は強い生物活性に重要な因子であるのに対しALS法はそれに加えて  $I(2-OR)$  及び  $I(R1)$  の寄与をも示している。共通データ種に関しての役割は両方法で同じであるが、ALS法は変数  $\Delta RI$  と  $\Sigma\pi$  を別々に扱うため、 $\Sigma\pi$  に関する係数を比較することができない。結論として、生物活性を決める主なデータ種は両方法で一致していると結論できる。

表5-10. アリルアクリロイルピペラジン<sup>a</sup>におけるALS型学習パターン使用時の再構築学習法の結合荷重行列と出力値

A <sup>b</sup>													
1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-0.001	0.539	0.001	0.001	0.000	0.000	0.000	0.672	-0.395	0.000	0.000	-0.001	-0.999
2	0.000	0.001	0.000	0.667	0.000	-0.001	0.000	-0.001	0.381	0.000	-0.001	0.000	0.335
3	0.000	-0.001	0.001	-0.805	0.000	0.000	0.000	-0.089	-0.239	0.000	0.000	-0.001	-0.443
4	0.000	0.001	0.000	-0.001	0.000	0.000	0.000	0.001	0.018	0.000	0.000	0.000	-0.376
5	0.000	0.004	0.001	-0.001	0.000	0.000	0.000	-0.001	-0.016	0.000	0.000	0.000	0.000
6	0.000	-0.301	0.001	-0.001	0.000	0.000	0.000	-0.028	0.188	0.000	0.000	-0.001	-0.010
7	0.000	-0.193	0.080	0.000	0.000	0.000	0.000	-0.290	0.000	0.000	0.000	-0.001	0.635

B <sup>c</sup>													
1	2	3	4	C <sup>d</sup>									
compd				1	2	3	4	obs	compd				
1	2	3	4	1	2	3	4	obs	16	17	18	19	20
1	0.000	0.000	0.000	0.000	0.984	0.151	0.000	0.000	0.984	0.991	0.449	0.002	2
2	0.001	0.001	-0.162	-0.555	0.984	0.037	0.000	0.032	0.984	0.999	0.880	0.014	3
3	0.242	0.000	0.000	0.000	0.984	0.064	0.000	0.037	0.984	1.000	0.998	0.057	3
4	0.001	0.377	0.263	0.395	0.984	0.001	0.000	0.003	0.983	1.000	1.000	0.155	3
5	0.001	0.000	0.000	0.000	0.983	0.028	0.005	0.056	0.984	0.890	0.961	0.180	3
6	0.000	0.000	-0.001	0.000	0.984	0.000	0.000	0.000	0.984	0.999	0.999	0.042	3
7	0.000	0.000	-0.001	0.000	0.984	0.000	0.000	0.000	0.984	0.999	0.972	0.012	3
8	0.001	0.000	0.000	0.000	0.984	0.269	0.169	0.054	0.984	0.996	0.994	0.006	3
9	0.001	-0.487	-0.696	-0.276	0.984	0.004	0.000	0.000	0.983	1.000	1.000	0.213	3
10	0.000	0.001	0.000	-0.615	0.984	0.005	0.000	0.000	0.984	1.000	1.000	0.958	4
11	0.001	0.000	0.000	0.000	0.984	0.000	0.000	0.000	0.984	0.995	0.890	0.889	4
12	0.000	0.000	0.000	0.000	0.984	0.880	0.007	0.001	0.983	1.000	1.000	0.876	4
13	0.001	0.000	-0.001	0.000	0.984	0.996	0.415	0.006	0.984	1.000	1.000	0.942	4
14	0.001	0.665	0.998	0.327	0.984	0.991	0.365	0.168	0.983	1.000	1.000	0.800	4
15	0.984	0.952	0.262	0.047	0.984	0.952	0.262	0.047	0.983	1.000	1.000	0.800	4

<sup>a</sup>The sum of error (E) < 0.142. <sup>b</sup>The weight matrix between the first (column) and second (row) layers (scale factor = 0.0616). <sup>c</sup>The weight matrix between the second (column) and third (row) layers (scale factor = 0.0484). <sup>d</sup>Outputs.



表5-11. アリルアクリロイルルピペラジンの生理活性発現に重要なパスと  
入カパラメータの寄与

path	$\Delta RI$	$\pi$	$\sigma$	I(2-OR)	I(R1)	HB
2	-0.64	0.00	0.08	0.00	0.00	0.30
4	0.00	0.67	-0.81	0.00	0.00	0.00
14	0.40	0.00	0.24	-0.02	0.00	-0.19

## 結 語

パターン認識ではパターン群の非線形分離機能が要求され、それを実現するためにニューラルネットワークは3層以上の階層的構造が必要となる。種々のニューラルネットワークのうちこの条件を満足し、有効な学習則を持ちかつ高速処理が期待出来るものはパーセプトロン型ニューラルネットワークである。

本論文では初めにパーセプトロン型ニューラルネットワークの学習則を解析し、その性質を検討した。すなわちニューラルネットワークで扱うデータは0～1区間の実数値に限定されるため、この区間に薬物の構造活性データをスケールしなければならない。学習データの数が少なくかつスケーリングについて対称性を満足していないときには、ニューラルネットワークの分類に偏りが見られた。この偏りは修正可能であるが、実際に薬物の構造活性相関に適用する場合、学習データの性質により、その偏りは軽微なことを示した。

構造活性相関への予備実験として、データに曖昧さの少ないノルボルネン誘導体の $^{13}\text{C}$ -NMR ケミカルシフト値と endo/exo 異性体の関係を用いて学習および予測能力を検討した。その結果ケミカルシフト値と endo/exo の関係を100%正しく学習することができ、未学習データについてもほぼ正しく（正答率94%）endo/exo 異性体を予測し、学習機械法およびクラスター分析法の85%を超えた。

この方法をマイトマイシン誘導体の制ガン活性、アリルアクロイルピペラジン誘導体の抗高血圧活性と分子構造データの関係に適用した。マイトマイシン誘導体で100%、アリルアクロイルピペラジン誘導体で90%の高率で相関関係を正しく認識できた。これらの結果は、現在もっとも信頼性の高いと言われている適応最小2乗法（ALS法）のマイトマイシン誘導体に対する94%、アリルアクロイルピペラ

ジン誘導体に対する62~76%を凌ぐものである。<sup>18, 23)</sup> さらにマイトマイシン誘導体の一部のデータを除外してネットワークの学習を行い、そのネットワークを用いて除外した未知データの生理活性強度を予測したところ、5段階分類において75%の正答率で生理活性等級が予測できた。

定量的な構造活性相関を求めるために、パーセプトロン型ニューラルネットワークの出力層のニューロンの数を一つにし、0~1の連続値をとらせた新型のニューラルネットワークを提案した。このネットワークの動作を解析し、その動作は非線形重回帰分析 (nonlinear multiregression analysis) 法に一致することを示し、MR型ニューラルネットワークと命名した。

本法をカルボキノン誘導体の制ガン作用、およびベンゾジアゼピン誘導体の抗不安作用と分子構造データとの関係に適用した。分子構造パラメータと生理活性強度について、ニューラルネットワークは重回帰分析法に比べて分散値で1.3~3.8倍強く関連性をつけることができた。

leave-n-out法 (n=1~10) により推測機能を調べたところ、データ中に矛盾の少ない場合にはニューラルネットワークは重回帰分析法よりも正しい推測を行う。しかし、データに自己矛盾が多い場合は、ネットワークが矛盾自体を規則として学習に取り込むため、未学習データに対する推測はデータの線形性を仮定した従来の重回帰分析法に及ばないことがある。前者の場合がカルボキノン誘導体で、後者はベンゾジアゼピン誘導体で見いだされた。

データに実験誤差が多い場合、重回帰分析法では決定係数が低く構造活性相関の結果が信頼出来ない。そこでニューラルネットワークの特性を用いて高い決定係数を保つ相関計算法を考察した。

重回帰分析では因果関係を示す係数として偏相関係数を用いる。MR型ニューラルネットワークの処理は重回帰計算の非線形拡張であるため、偏相関係数をニューラルネットワークで計算するアルゴリズムをつくることが出来る。ニューラルネットワークを十分に学習させて教師データとネットワークの出力を完全に一致させ、かつ非線形パラメータを制御して線形ネットワーク動作に近づけると、重回帰分析から計算される偏相関係数とほとんど一致する値が得られることがわかった。この方法は重回帰分析法を拡張したもので理論的には明快であるが、数値的なゆらぎによる係数の再現性を保つためには、通常の学習よりネットワークの学習を10倍以上厳密に行わなければならない、実用的には困難である。

そこでニューラルネットワークに入力されるデータの微小変化を追跡し、出力と入力データの比（相関指数）を求める新しい方法を考案した。この方法に不飽和型のニューロン動作関数を導入し、学習方程式に「ニューロンの疲労」という概念を取り入れた。この方法は前者に比べ、20倍程度高速に計算でき、さらに学習回数の異なりによる計算結果のゆらぎが無く、再現性に優れている。このように計算された入力、出力間の相関指数は重回帰分析計算の偏相関係数と正しく対応する。すなわちニューラルネットワークを用いた入力ー出力間の相関計算の新しいインデックスを導入した。

パーセプトロン型ニューラルネットワークに、学習という自己組織構築状態と同時に形成した自己組織の一部を消去する（忘却）状態をも取り入れることができる。この2つの状態を混合し、学習も忘却も成立しない状態をニューロシミュレータで実現する。すなわち、不十分に学習したニューラルネットワークに交互に学習と忘却の両状態を取らせると、シナプス強化則に従いニューラルネットワーク内のニューロン間の大部分の結合は淘汰され、重要な結合だけが残りがつ強化される。この結果、活性強度等出力を決定する入力層のニューロンを定めるこ

とができることがわかった。この方法を再構築学習と命名した。

再構築学習法を、ノルボルネン誘導体の $^{13}\text{C}$ -NMR ケミカルシフト値と置換基の endo/exo 配向性の関係に適用したところ、置換基から最も近い位置にある2つの炭素原子に関するデータを受け入れているニューロンに関する結合が置換基の配向性に関係していることが示された。またマイトマイシンおよびアクリルアクロイルピペラジン誘導体の構造活性データに適用し、薬理作用発現の構造因子を抽出することができた。

これらの研究において、ニューラルネットワークは従来方法と比較してはるかに優れた方法であることを確認した。さらに、ニューラルネットワークには曖昧なデータを扱えることとノンパラメトリックであることの利点がある。難点は学習に時間がかかることである。これはニューロコンピュータを通常のコンピュータでシミュレートしているためで、ニューロチップ等の利用でこの問題は解決されるものと思われる。

## 参考文献

- 1) W. S. McCulloch and W. Pitts, Bull. Math. Biophysics, 5, 115 (1943).
- 2) D. O. Hebb, "The Organization of Behavior," Wiley, New York, 1949.
- 3) F. Rosenblatt, "Principles of Neurodynamics and the Theory of Brain Mechanics," Spartan Books, Rockelle Park, 1961.
- 4) H. D. Block, Rev. Mod. Phys., 34, 123 (1962).
- 5) M. Minsky and S. Papert, "Perceptron - An Essay in Computational Geometry," MIT Press, Cambridge, Ma., 1969.
- 6) 中野 馨, "アソシアトロンとその応用 - 連想記憶装置に関する研究," 電子通信学会インフォメーション理論研究会資料, IT69-27, 1969;  
K. Nakano and J. Nagumo, Proc. 2nd Int. Joint Conf. on Artificial Intelligence, 101, 1971;  
K. Nakano, "Learning process in a model of associative memory in Pattern Recognition and Machine Learning," K. S. Fu, Ed., Plenum Press, New York, 172, 1971.
- 7) 上坂吉則, 尾関和彦, "連想記憶の2, 3の性質," 信学論, 55-D, 323, 1972.
- 8) J. A. Anderson, Math. Biosciences, 14, 197 (1972).
- 9) T. Kohonen, IEEE Trans., C-21, 353 (1972).
- 10) K. Fukushima, Biol. Cybern., 20, 121 (1975); *ibid.*, 36, 193 (1980).
- 11) 甘利俊一, "神経回路網の数理," 産業図書, 1978.
- 12) C. von der Malsburg, Kybernetik, 14, 85 (1973).
- 13) J. J. Hopfield, Proc. Natl. Sci. U.S.A., 79, 2554 (1982).

- 14) D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, *Cognitive Sciences*, 9, 147 (1985); G. E. Hinton and T. J. Sejnowski, "Parallel Distributed Processing," Vol. 1, ed. by D. E. Rumelhart and J. L. McClelland, MIT Press, Cambridge, Ma., 1986. Chap. 7.
- 15) D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Nature*, 323, 533 (1986); D. E. Rumelhart, J. L. McClelland and PDP Research Group, "Parallel Distributed Processing", MIT Press, Cambridge, Ma., 1986.
- 16) T. J. Sejnowski and C. R. Rosenberg, "NETtalk: A Parallel Network That Learns to Read Aloud," The Johns Hopkins University Electrical Engineering and Computer Science Technical Report," JHU/EECS-86/01, 32, 1986 and reported in "Neurocomputing: Foundations of Research", ed. by J. A. Anderson and E. Rosenfeld, MIT Press, Cambridge, Ma., 1988.
- 18) I. Moriguchi, "In Structure-Activity Relationship Quantitative Approaches"; T. Fujita, Ed., Nankodo: Tokyo, Japan, 1986; chapter 9.
- 19) T. Aoyama, Y. Suzuki, and H. Ichikawa, *J. Med. Chem.*, 33, 2583(1990).
- 20) T. Aoyama and H. Ichikawa, *Chem. Pharm. Bull.*, 39, 358(1991).
- 21) S. Sasaki, H. Abe, Y. Takahashi, T. Takayama, and Y. Miyashita, "Introduction to Pattern Recognition for Chemists", Tokyo Kagaku Dojin, Tokyo, Japan, Chap. 1(1984).
- 22) T. Sekiya, H. Hirayama, S. Hata, S. Mizogami, M. Hanazawa, S. Yamada, *J. Med. Chem.*, 26, 411(1983).
- 23) T. Sekiya, "In Structure-Activity Relationship and Drug Design", T. Fujita, Ed., Kagaku-dojin, Kyoto, Japan, 1986, Chapter 9.

- 24) T. Aoyama, Y. Suzuki, and H. Ichikawa, Chem. Pharm. Bull., 37, 2558 (1989).
- 25) T. Aoyama, Y. Suzuki, and H. Ichikawa, J. Med. Chem., 33, 905(1990).
- 26) H. Nakao, M. Arakawa, T. Nakamura, M. Fukushima, Chem. Pharm. Bull., 20, 1962(1972).
- 27) H. Nakao and M. Arakawa, Chem. Pharm. Bull., 20, 1968(1972).
- 28) C. Hansch, R. M. Muir, P. Maloney, E. Geiger, M. Streich, J. Am. Chem. Soc., 86, 2817(1963).
- 29) C. Hansch and T. Fujita, J. Am. Chem. Soc., 86, 1616(1964).
- 30) C. Hansch, J. Med. Chem., 19, 1(1976).
- 31) M. Yoshimoto, H. Miyazawa, H. Nakao, K. Shinkai, M. Arakawa, J. Med. Chem., 22, 491(1979).
- 32) L. O. Randall, W. Schallek, G. A. Heise, E. F. Eith, R. E. Bagdon, J. Pharmacol. Exp. Ther. 129, 163(1960).
- 33) L. H. Sternbach, L. O. Randall, R. Banziger, H. Lehr, In Medicinal Research Series 2. Drugs Affecting the Central Nervous System, A. Burger, Ed., Merrell Dekker, Inc., New York, 1968, p.237.
- 34) T. Kubota, M. Yamakawa, H. Terada, M. Yoshimoto, "In Structure-Activity Relationships-Quantitative Approaches," the QSAR Research Group, Ed. Kagaku no Ryoiki Supl. Ed., No.122(1979).
- 35) J. S. Driscoll, G. F. Hazard Jr., H. B. Wood Jr., A. Goldin, Cancer Chemother. Rep. Part 2, 4,1(1976).
- 36) A. Kharn, J. S. Driscoll, J. Med. Chem., 19, 313(1976).



- 37) F. Chou, A. H. Khan, J. S. Driscoll, J. Med. Chem., 19, 1302(1976).  
 38) T. Aoyama and H. Ichikawa, Chem. Pharm. Bull., 39,1222(1991).  
 39) T. Aoyama and H. Ichikawa, Chem. Pharm. Bull., 39, 372(1991).  
 40) K. Funahashi, Neural Networks, 2, 183(1990).  
 41) 計算した薬物の種類, 誘導体数, 3層のニューラルネットワークで正しく分類できた比率は次のようである。

化合物名	活性	誘導体数	比率
マイトマイシン	抗ガン	16	100%
N-ピペリジノ	鎮痛	26	89
プロピルカルボン酸アミド	抗炎症	26	81
アリルアクリロイル			
ピペラジニル	抗高血圧	29	90
モルホリノカルボン酸	鎮痛	38	95
アミド	抗炎症	38	97

- 42) M. Kayama, S.Abe, H.Takenaga and Y.Morooka, Neuro Nimes 90, Nimes, France, Nov. 1990.  
 43) M.S.Klassen and Y.H.Pao, IEEE Proceedings of 2nd Annual International Conference on Neural Networks, San Diego, CA., June 1988.

## 謝 辞

本研究は星薬科大学市川 紘助教授及び鈴木雄二博士（現フナコシ薬品）の御指導，御支援の基に達成されたものであります。

市川 紘助教授は構造活性相関の問題点を初めとする本研究全般についてご指導され，鈴木雄二博士は分子構造及び薬理活性データの意味についてご教授されました。

東北大学薬学部山中 宏教授は，本研究を支援され，本論文の作成を指導され，さらに東北大学大学院薬学研究科に学位申請の機会を与えてくださいました。

ここに深く感謝致します。

この論文に述べられている計算結果は日立コンピュータエンジニアリングの所有するパーソナルコンピュータ PC-9801/RA21 + INMOS T-800 ボード，MS-DOS，及び 3L PARALLEL-FORTRAN を用いて行ったものであります。

付録1. MR型ニューラルネットワークの関数fitting機能の限界

第2, 3層の各々の1個のニューロンの動作関数が直交関数系 $e_\nu$ , 線形関数のニューラルネットワークにおいて, 第1, 2層間の結合荷重が $\delta_{ij}w_{ij}$ となっている, 一種のMR型ニューラルネットワークを考える.  $\delta$ はデルタ関数である. この型のニューラルネットワークをfunctional network<sup>43)</sup>と呼ぶ. このネットワークで或る関数 $h(x)$ を, 近似する場合,  $h(x) \sim \sum w_{jk}e_\nu \equiv g(x)$ , なる関係が成立している. ここで $w_{jk}$ は第2, 3層間の結合荷重である. これを,  $F(h(x)) = g(x)$ と書く. 次に $h(x) - g(x)$ なる関数を, 同じニューロン動作関数を有する別の2つのfunctional networkで近似するとき高々,

$$F(h(x) - g(x)) = g(x) - 0 = g(x), \quad A-1$$

である. 従って $\{e_\nu\}$ が同じニューロン系でニューロン数を増加させても関数近似能力は向上しない.

一般のMR型ニューラルネットワークは, 直交関数系 $\{e_\nu\}$ をベクトル $\eta$ で変換して,  $f = \eta e_\nu$ とし, この $f$ をシグモイド関数とすると得られる. ただし第1, 2層間の結合荷重は $\delta_{ij}w_{ij} \rightarrow w_{ij}$ となっているから, その自由度の向上で,

$$\Omega(h(x) - g(x)) = g'(x) - 0 = g'(x), \quad A-2$$

$$\int \{g'(x)\}^2 dx < \int \{g(x)\}^2 dx, \quad A-3$$

となる. ここで $\Omega$ はMR型ニューラルネットワークによる変換作用をあらわす.

しかし数値計算上は $w_{ij}$ の取り得る範囲が数表現の制限によって限定されるので,

$\int \{g'(x)\}^2 dx > 0$ である. 再度関数の差をfitさせようとしても,

$$\Omega(h(x) - g(x) - g'(x)) = g'(x), \quad A-4$$

となる. 従って任意の関数を完全にfittingすることは出来ない. 以上の議論で,

$\Omega(h(x) - g(x)), \dots$ という変換は $w_{ij}$ という係数行列の限定範囲内の自由度によって, 新しい直交関数系 $\{e_\nu\}$ を作り出すだけであるから, 極めて非効率的な変換であることが予想できる. 経験的には, MR型ネットワークの関数fitting結果と

目的関数の差を計算して、その差を新たにMR型ニューラルネットワークでfittingしても、誤差関数E値は殆ど小さくならなかった(0.7倍程度)。

このような性質があるため、ニューラルネットワークは剛性をもち、それによって、第2層が必要最少限のニューロン数でなくとも、あまり極端に多くなければ、誤差が含まれるデータの中から重要な関係を抽出できると考えられる。

#### 付録2. 4次元以上の4分類系

3次元4分類系を、 $\{ (C, Q^{(3)} C) \}$  , と表現するとき、4次元は、

$$[ \{ (C, Q^{(3)} C) \}_0, Q^{(4)} \{ (C, Q^{(3)} C) \}_0 ],$$

である。ここで、

$Q^{(4)} = (Q^{(3)}, +1)$  , である。 $+1$ は4次元になって新しく増加した座標について1とすることを意味する。故に、

$(+1) (C, Q^{(3)} C) \}_0 = (C, Q^{(3)} C) \}_1$  , である。以下 $Q^{(n)}$ 演算子を漸化式で作れば良い。

#### 付録3. 微分バックプロパゲーション法

ニューロン間の結合荷重の変化に対して、ニューラルネットワークの入出力に関する偏微分係数は、

$$\text{各層について, } \delta y_j = W_{ij} \delta x_i, \quad \delta O_j = f'(y_j) \delta y_j \quad A-5$$

であるから、2層のネットワークでは、

$$\delta O_j / \delta x_i = f'(y_j) W_{ij}, \quad A-6$$

3層では、

$$\delta O_j / \delta x_i = \sum_k f'(y_k) V_{ik} g'(z_j) W_{kj} \quad A-7$$

である。ただし $f'(y_k)$  ,  $g'(z_j)$ はそれぞれ2, 3層のニューロンの動作関数の微分関数である。 $V_{ik}$ は1, 2層間の、 $W_{kj}$ は2, 3層間の結合荷重である。

ニューロン間の結合荷重の変化に対して式A-7は次のように変化する。

$V_{IK}$ の微小変化  $\varepsilon$  に対して、 $y_K \rightarrow y_K + \xi$  と変化するとすると、

$$\xi = x_I \varepsilon,$$

$$f'(y_K + \xi) = f'(y_K) + f''(y_K) \xi \quad A-8$$

$\xi$  という変化に対し第2層のニューロンの出力値の変化は、

$f(y_K) + f'(y_K) \xi$ 。これは $W_{KJ}$ を介して $g'(z_J)$ を変化させる。

$$z_J + \eta = W_{KJ} \{f(y_K) + f'(y_K) \xi\},$$

$$\eta = W_{KJ} f'(y_K) \xi = W_{KJ} f'(y_K) x_I \varepsilon,$$

$$g'(z_J + \eta) = g'(z_J) + g''(z_J) \eta \quad A-9$$

故に式A-7の  $\varepsilon$  に対する変化 $\Delta_{IK}$ は、

$$\begin{aligned} & (\delta O_J / \delta x_I) + \Delta_{IK} \\ &= \sum_{k \neq K} g'(z_J) W_{KJ} f'(y_K) V_{Ik} \\ &+ [g'(z_J) + g''(z_J) \eta] W_{KJ} [f'(y_K) + f''(y_K) \xi] (V_{IK} + \varepsilon) \end{aligned} \quad A-10$$

となる。ここで、

$$\begin{aligned} \Delta_{IK} &= W_{KJ} V_{IK} [g''(z_J) \eta f'(y_K) + g'(z_J) f''(y_K) \xi] \\ &+ W_{KJ} g'(z_J) f'(y_K) \varepsilon \end{aligned} \quad A-11$$

$$\begin{aligned} \Delta_{IK} &= W_{KJ} [V_{IK} [g''(z_J) W_{KJ} \{f'(y_K)\}^2 + g'(z_J) f''(y_K)] x_I \\ &+ g'(z_J) f'(y_K)] \varepsilon \end{aligned}$$

A-12

一方、 $W_{KJ}$ の微小変化  $\varepsilon$  に対して、 $z_J \rightarrow z_J + \eta$  と変化するとすると、

$$\eta = f(y_K) \varepsilon,$$

$$g'(z_J + \eta) = g'(z_J) + g''(z_J) \eta$$

$$\begin{aligned}
& (\delta O_J / \delta x_I) + \Delta_{KJ} \\
& = \sum_{k \neq K} g'(z_J) W_{KJ} f'(y_K) V_{IK} \\
& + [g'(z_J) + g''(z_J) \eta] (W_{KJ} + \epsilon) f'(y_K) V_{IK} \quad A-13
\end{aligned}$$

となる。ここで、

$$\begin{aligned}
\Delta_{KJ} &= [g'(z_J) \epsilon + g''(z_J) \eta W_{KJ}] f'(y_K) V_{IK}, \\
\Delta_{KJ} &= [g'(z_J) + g''(z_J) f(y_K) W_{KJ}] f'(y_K) V_{IK} \epsilon \quad A-14
\end{aligned}$$

である。入力、教師データの他に偏微分係数が与えられたとき、

$$E = \sum_J (O_J - t_J)^2 + \lambda \sum_{(I, J)} \{ (\delta O_J / \delta x_I) - \rho_{IJ} \}^n \quad A-15$$

を結合荷重  $V$ ,  $W$  で偏微分し、その式を  $=0$  と置いて学習方程式を作ることができる。

ここで、 $\lambda$ ,  $n$  は定数、 $\rho_{IJ}$  は偏微分係数の期待値である。

$n=1$  のとき式 A-15 の  $\lambda$  項は、

$$\begin{aligned}
q_{IJ} &= (\delta O_J / \delta x_I) - \rho_{IJ}, \\
\delta V_{IK} &= -\lambda L [\sum q_{IJ} \Delta_{IK}^{-1} \{1 - d(\Delta_{IK})\}], \quad A-16
\end{aligned}$$

$$\delta W_{KJ} = -\lambda L [\sum q_{IJ} \Delta_{KJ}^{-1} \{1 - d(\Delta_{KJ})\}], \quad A-17$$

である。ここで  $d(x)$  は  $x=0$  のとき 1, それ以外のとき 0 となる関数、

$L[x]$  は  $|x| \geq \theta$  のとき  $\text{sgn}(\theta, x)$ , それ以外のとき  $x$  となる関数である。

式 A-15 は、

$$\rho_{IJ}^{(\mu)} = M^{-1} \sum_{(J, I)} (\delta O_J / \delta x_I)^{(\mu-1)} \quad A-18$$

の様に  $M$  個のデータについて平均し、次回の学習の偏微分係数の期待値  $\rho_{IJ}^{(\mu)}$  を巡回操作の中で計算する形に変形できる。また SAR の適用では相隣合った等級間で、

$\rho_{IJ}^{(\mu)} = -\rho'_{IJ}^{(\mu)}$  の様にして、ニューラルネットワークの非対称性分類問題をなるべく極小化するような条件を導入する方向に変形出来る。

実際の学習は次の様に行なう。

1) バックプロパゲーション学習で $E \sim 0.001$ までニューラルネットワークを学習する。

2) 式A-16, 17でニューロン間結合荷重を補正する。E値は大きくなる。

定数 $\lambda$ ,  $\theta$ はそれぞれ0.001, 0.01程度にとる。

3) Eの値をもとの値よりも僅かに小さくなるようにバックプロパゲーション学習を行なう。

4) 2, 3の操作を $E \rightarrow 0$ まで繰り返す。Eが小さくならない時 $\lambda$ を小さくするが、

$\lambda \rightarrow 0$ となるときは第2層のニューロンが少ない。

この方法はバックプロパゲーション学習による結合荷重の補正部と、式A-15の $\lambda$ 項補正部が独立した操作になっているため、条件の導入が容易に出来る。

#### 付録4. スケーリング, 等級分類表現における問題点

入力データ種個々について独立した2.2.2節のようなスケーリングは、ニューロンの定義域を全部使用するように出来るので数値計算精度上は好都合であるが、 $\phi$ 線分の交点を変化させるためニューラルネットワークの推測機能に影響を与える。一般的には $\phi$ 線分の交点を不変とする様な入力データ種全てに依存するスケール法が必要である。但し薬物の等級予測の場合、3.3.2節の議論より $\phi$ 線分の交点が存在しないので、スケーリングについて問題にならないと思われる。

等級分類表現では $m$ 個のカテゴリーを $m$ 個のベクトル $v_k = (\dots, \delta_{ik}, \dots)$ で表現することが一般的に行われている。 $\delta_{ik}$ はデルタ関数、 $k$ はカテゴリー番号である。そしてニューラルネットワークの出力を、 $S_k = \sum (O_i - v_{ik})^2$ で計算し、最小の $S_k$ を与える $k$ をネットワークが認識したカテゴリーとしている。

この方法は、任意の $v_k$ ベクトルを一点とする空間で、各 $v_k$ 点の周辺で等価なネットワーク出力分布でなければ正しくない。3.1節の例でも判るように等価なネットワーク出力分布が期待出来るのは、 $v_k$ 点の周りの小さい領域である。

付録5. ニューラルネットワークに与えるデータの変動に対する結果の変化

薬理活性データには無視できない量の測定誤差が含まれている。そのとき薬理活性データをニューラルネットワークによって解析することの妥当性を議論する。

ニューラルネットワークに入力するデータ及び教師データの変動を次の様に表示する。

$$\begin{aligned}x_i^* &= x_i + \mu_i, \\t_i^* &= t_i + \nu_i\end{aligned}\tag{A-19}$$

ここで  $x^*$ ,  $t^*$  は実際の測定値,  $x$ ,  $t$  は誤差無しの理想条件で測定するとき得られると予想される真の測定値である。 $\mu$ ,  $\nu$  は擾乱項である。

この項による2-9, 10式の変化は次の様である。ネットワークは2層とする。

$$\delta w_{ij} = -d_j x_i^* \varepsilon = -d_j x_i \varepsilon - d_j \mu_i \varepsilon,\tag{A-20}$$

$$\begin{aligned}d_j &= (O_j - t_j^*) f'(y_j^*) \\&= (O_j - t_j) \{f'(y_j) + f''(\sum w_{ij} \mu_i)\} - \nu_j f'(y_j).\end{aligned}$$

従って,
$$\tag{A-21}$$

$$\begin{aligned}\delta w_{ij} &= (O_j - t_j) f'(y_j) \\&\quad - \{\nu_j f'(y_j) - d_j \mu_i \varepsilon - (O_j - t_j) f''(\sum w_{ij} \mu_i)\}.\end{aligned}$$

{ } 内の項が学習方程式に加えられた擾乱である。
$$\tag{A-22}$$

$$\begin{aligned}&\{\nu_j f'(y_j) - d_j \mu_i \varepsilon - (O_j - t_j) f''(\sum w_{ij} \mu_i)\} \\&= \nu_j f'(y_j) - (O_j - t_j) f'(y_j) \mu_i \varepsilon \\&\quad - (O_j - t_j) f''(\sum w_{ij} \mu_i),\end{aligned}\tag{A-23}$$

$(O_j - t_j)$  は学習が完了すると0になるから,  $\{ \} \rightarrow f'(y_j) \nu_j$  である。  
 $f$  がシグモイド関数ならば,

$$Z = \nu_j / 4,\tag{A-24}$$

以下の不確実性が学習方程式に存在することになる。



3層の場合，notationを次のようにする。

$x_i$  : 真の入力データ，

$y_k$  : 真の入力データ時の第2層のニューロンに入力される値，

$z_j$  : 真の入力データ時の第3層のニューロンに入力される値，

$t_j$  : 真の教師データ，

$x^{\#}_i$  : 実際の入力データ，

$y^{\#}_k$  : 実際の入力データ時の第2層のニューロンに入力される値，

$z^{\#}_j$  : 実際の入力データ時の第3層のニューロンに入力される値，

$t^{\#}_j$  : 実際の教師データ，

$v_{ik}$  : 1, 2層間の結合荷重，

$w_{kj}$  : 2, 3層間の結合荷重，

$f$  : 第2層のニューロンの動作関数，

$g$  : 第3層のニューロンの動作関数。

次の関係が存在する。

$$x^{\#}_i = x_i + \mu_i,$$

$$t^{\#}_j = t_j + \nu_j, \quad A-19$$

$$y^{\#}_k = y_k + \sum v_{ik} (x_i + \mu_i), \quad A-25$$

$$z^{\#}_j = z_j + \sum w_{kj} g(\sum v_{ik} \mu_i). \quad A-26$$

学習方程式の揺らぎは，2, 3層間で，

$$\delta w_{kj} = -d_j^{(2,3)} y^{\#}_k \epsilon, \quad A-27$$

$$d_j^{(2,3)} = (O_j - t^{\#}_j) f'(z^{\#}_j), \quad A-28$$

である。1, 2層間では，

$$\delta v_{ik} = -d_k^{(1,2)} x^{\#}_i \epsilon, \quad A-29$$

$$d_k^{(1,2)} = \{\sum w_{kj} d_j^{(2,3)}\} g'(y^{\#}_k), \quad A-30$$

である。

A-27から,

$$\begin{aligned}
 (-\varepsilon^{-1}) \delta w_{kj} &= (O_j - t_j - v_j) \\
 &\times \{f'(z_j) + f''(\sum w_{kj} g(\sum v_{ik} \mu_i))\} \{y_k + \sum v_{ik} (x_i + \mu_i)\} \\
 &\sim (O_j - t_j) f'(z_j) y_k - v_j f'(z_j) y_k \\
 &\quad + (O_j - t_j) [f''(\sum w_{kj} g(\sum v_{ik} \mu_i)) y_k \\
 &\quad + f'(z_j) \sum v_{ik} (x_i + \mu_i)].
 \end{aligned} \tag{A-31}$$

学習完了時

$$\delta w_{kj} = v_j f'(z_j) y_k \varepsilon \leq v_j y_k \varepsilon / 4 \equiv Z, \tag{A-32}$$

である。係数1/4はシグモイド関数のときに限る。

A-29から,

$$\begin{aligned}
 (-\varepsilon^{-1}) \delta v_{ik} &= [\sum w_{kj} (O_j - t_j - v_j) \\
 &\times \{f'(z_j) + f''(\sum w_{kj} g(\sum v_{ik} \mu_i))\}] \\
 &\times \{g'(y_k) + g''(\sum v_{ik} (x_i + \mu_i))\} (x_i + \mu_i) \\
 &\sim [\sum w_{kj} (O_j - t_j) f'(z_j)] g'(y_k) x_i \\
 &\quad - [\sum w_{kj} v_j f'(z_j)] g'(y_k) x_i \\
 &\quad + [\sum w_{kj} (O_j - t_j) f''(\sum w_{kj} g(\sum v_{ik} \mu_i))] g'(y_k) x_i \\
 &\quad + [\sum w_{kj} (O_j - t_j) f'(z_j)] g''(\sum v_{ik} (x_i + \mu_i)) x_i \\
 &\quad + [\sum w_{kj} (O_j - t_j) f'(z_j)] g'(y_k) \mu_i
 \end{aligned} \tag{A-33}$$

この式は学習完了時,

$$\begin{aligned}
 \delta v_{ik} &= [\sum w_{kj} v_j f'(z_j)] g'(y_k) x_i \varepsilon \\
 &\leq [\sum w_{kj} v_j] x_i \varepsilon / 16 \equiv Z,
 \end{aligned} \tag{A-34}$$

である。係数1/16はシグモイド関数のときに限る。

A-24, 32, 34から学習方程式2-9, 10, 11は教師データの変化に対して1次の依存関係がある。入力データの変化には関係しない。重回帰分析法では共に1次の依存関係であるが、同じ1次依存関係でも係数は1である。

従ってパーセプトロン型ニューラルネットワーク法は重回帰分析に比べて測定誤差に強い方法である。

教師データの誤差に対しては、3.1.2節に示したように $\sum (O_j - t_j)^2 \sim 0.15$ でネットワークの機能は実現されるので、それより1オーダー小さい擾乱ならば学習は有意と思われる。従って、 $v_j \sim 0.49$ である。これは49%程度の測定誤差があってもパーセプトロン型ニューラルネットワークで有意な分類が可能であることを示す。

このことを次のような数値実験で確かめた。

1. 入力データを  $y = x$ ,  $(0, 1)$  区間からサンプリングした値,
2. 教師データに  $y = \sin(x)$ ,  $(0, 2\pi)$  区間からサンプリングした値と  $(-a, a)$  区間の一様乱数の符号を1サンプリング毎に変えたものを加えた値,

をMR型ニューラルネットワークで学習させた。サンプリング数は10000である。学習完了後ネットワークに  $y = x$  を入力し、出力波形を調べた。その結果、 $a = 0.5 \sim 1$  ならばニューラルネットワークの出力は正弦波であるが、 $a = 2$  では正弦波の再現は不可能であった。この結果は50~100%程度の誤差ならば或る程度の有意な処理がニューラルネットワークで可能なことを示し、上記の理論的考察を支持している。

多くの測定誤差が含まれる系の学習では、学習完了時  $\delta E = 0$  である必要はなく、 $\delta E \sim Z$  で良い。すなわち、測定誤差から推測される  $Z$  以下に学習方程式を収束させても無意味である。それよりもニューロンの動作関数に線形性を導入して、学習の収束点を  $Z$  程度に止め、ネットワークの汎化作用を強くすべきである。強い非線形ニューロン動作関数では複数のデータ間で  $f'(y_j)$  項の相違が大きくなり、測定誤差に対する結果の変動が不均一になる傾向が強調される。これは望ましくない。